Classification of Cars in a Multiplicative Rating Model using Recursive Credibility Estimation

—Theory and Application

by

Arne Eyland

Preface

The present paper is a revised version of the author's thesis written in order to obtain the degree of cand.scient. at the University of Oslo, Institute of Mathematics, Department of Statistics and Insurance Mathematics. I am indebted to both the external examiner of my thesis, Professor André Dubey, and my teaching supervisor, Professor Bjørn Sundt, for giving valuable comments on the thesis. These comments have made it possible for me to produce this revised version.

The research leading to the thesis was started in October 1988, and the approach to the problem was given to me by Bjørn Sundt. The practical part of the thesis originated from a summer position I held in the actuarial department of the Norwegian insurance company Storebrand during the summer of 1990.

Several persons have helped me during the preparation of the thesis.

I am particularly grateful to Bjørn Sundt. He has gone through earlier versions of the thesis and several improvements are results of his comments. We have had fruitful discussions on numerous occasions about topics concerning the thesis. He has been a valuable source of inspiration to me and he has stretched far beyond what I could require of a supervisor. Thanks to Bjørn Sundt encouraging me, I have presented parts of an earlier version of the thesis at the 23rd ASTIN Colloquium in Stockholm.

I will also like to thank the staff in the actuarial department of Storebrand (in January 1991 the two Norwegian insurance companies UNI and Storebrand merged and became UNI Storebrand) for their excellent cooperation. They have given me access to their vast data material and thus made it possible for me to check the performance of one of the models proposed in this paper. This access was not given to me only when I held the summer position in the company, but actually until I finished my studies. I am grateful to them for letting me include some of these data in the present paper. They have also been very patient with me and obliging whenever I needed help in solving a practical problem.

Thorbjørn Narvestad from the motor insurance department in UNI Storebrand has been very willing to help me whenever I needed an expert opinion concerning cars. He has also selected the car models to be included in the numerical examples in this paper and gathered the technical data about these car models.

Ola Roll and Stein Barholt from the same department have also been a help to me.

November 20, 1993 — Arne Eyland

Contents

Chapter 1

Introduction

In the present paper we are going to study ways of determining a classification of cars according to the risk they constitute in a motor insurance portfolio. We consider both vehicle damage and (third party) liability covers and confine ourselves to passenger cars, estate cars and vans (with carrying capacity up to 1 200 kilograms).

We shall work within a multiplicative rating model used in Storebrand. This rating model consists of factors for district, mileage, car model, bonus/malus etc. In this paper we will concentrate on the factor for car model and treat the other factors as given. Since a larger risk should have a higher premium, we have that a large risk should have a high value of the factor for the car model or, equivalently, be in a high risk class. In vehicle damage insurance, Storebrand used 65 risk classes numbered from 30 to 94. The factor given to class c was 1.04^{c-30}. In liability insurance they used 6 classes numbered from 1 to 6 with factors 0.75, 1.00, 1.07, 1.13, 1.33 and 1.50, respectively. In this paper we will adopt these factors and classes to match. In other words we are in the same situation that has previously been studied by SUNDT(1987a). For a more extensive discussion of the rating model we refer to SUNDT(1987a, pp. 41–43) and SUNDT(1991b, Chapter 8).

We will not give any further description of what we mean by a car model. It should be chosen in each particular situation in an appropriate way. By a car model we could e.g. mean makes, models or variants of cars.

We want to classify both car models which are new (for which the risk volume is equal to zero) and car models which we have observed. The two cases will be treated simultaneously since the former can be regarded as a special case of the latter.

The portfolio consists of a number of individual policies, and to each policy there is associated a car model. Hence, to the most popular car models we may base the classification on a large number of observations. The number of policies may vary from year to year for a specific car model. Different policies are allowed to stay in the portfolio for a different number of years. We have observations from the portfolio during the past n years. The present author feels that these allowances are necessary for the models to be realistic. The drawback is that this results in a bothersome notation; we have to introduce a large number of definitions. To help the reader keep track of the various definitions we have at the end of the paper made an index containing a list of notations. This list contains most of the symbols used in the text and a reference to the page they were defined or used on for the first time. The classification of each car model is revised once a year. Assume that we are at the end of year n and want to estimate the risk level in year $n+1$ for all the car models in the rating structure using credibility estimation. For car models included in the portfolio at this point in time we have at hand the claim history of the associated individual policies. We are going to present models describing how to utilize this information and how to combine it with prior information about the risk level of the car model in year n and $n+1$.

We will develop models incorporating observations from several past years. By including observations from not only one year we will hopefully get a considerably better estimator of the risk level of each car model, and thus a more correct classification. This is an important aspect for an insurance company. Too high a classification for a car model will give too high a premium for that car model and the company may be outstripped by its competitors. On the other hand, too low a classification will give too low a premium and the company will lose money.

SUNDT(1987a) has derived models to deal with the same situation as in the present paper. But by using his models observations from different years will be given the same weight. This is slightly unfortunate. Since both car models and the community as a whole are under constant change we feel that observations from recent years are more relevant and thus should be given more weight than old ones. We will state models taking this into consideration by viewing the risk levels of a car model as a stochastic process developing over time, and allowing their distribution to change as time passes. In these models the credibility estimators from year to year will be calculated recursively and are thus easy to handle in practice. These are the same ideas that form the basis of SUNDT(1981). We will use a special case of one of his models on each car model.

Another approach to the extension of the models in SUNDT(1987a) by including more observations, has been done in SUNDT(1987b). But this approach lacks the formal model apparatus. A consequence of this is that the estimator can not in any sense be shown to be optimal. But under weak assumptions it can be shown to be better than the credibility estimator in SUNDT(1987a).

We are going to consider three models:

- A time-heterogeneous model (Chapter 5)
- A time-homogeneous model (Chapter 6)
- A time-heterogeneous model for two portfolios (Chapter 7)

In the first model the structural parameters are allowed to vary from year to year. In the second model the parameters are constant over time. As observable risk characteristics we use technical variables like price, weight and engine power. But for car models that are no longer for sale in Norway, the price will be unknown. This motivates the introduction of the third model. Here we divide the original portfolio into two sub-portfolios, each having its own technical variables. One sub-portfolio consists of car models still on the market in Norway. The other consists of car models no longer for sale in Norway (and therefore the price no longer exists). We now use the theory from the second model on each of the two sub-portfolios.

Notational conventions used throughout the paper, important results, and notions are stated in Chapter 2. In Chapter 3 we make use of the concept of linear sufficiency and give sufficient conditions for the credibility estimator to be linear sufficient. Hence we reduce the dimension of the above described estimation problem considerably, which is of great practical importance. Model assumptions and definitions of symbols used in Chapters 5 and 6 are given in Chapter 4. Our models are stated and analysed in Chapters 5–7. We study the connection between recursive credibility estimation and the (linear) Kalman filter in Chapter 8. The relations between our models and well-known models other than the Kalman filter are indicated in Chapter 9. It is not necessary to read these two chapters in order to understand the rest of the paper. The purpose of Chapters 8 and 9 are merely to put the models developed in this paper into a wider context. In Chapter 10 we undertake the task of estimating the structural parameters. The estimators are based on a sample of car models taken from the portfolio (possibly with an exception of the estimator of φ_j, to use the notation from Chapter 5). In Chapter 11 we make some preparations for a practical application of our models including derivation of an alternative estimator of φ_j based on a Poisson-assumption. An application of the model in Chapter 5 on real data, is given in Chapter 12. In Chapter 13 we discuss the pros and cons of the models proposed in this paper. Some topics which should be further investigated and other interesting subjects relevant to the models presented in this paper, are given in Chapter 14. Identities frequently referred to in the text are derived in Appendix A. At the end of the thesis we have included an index of symbols as mentioned above.

Chapter 2

Preliminaries

2.1 Introduction

In the present chapter we will give some notational conventions, define some concepts frequently referred to in this paper, state a well known result in credibility theory, and give an important property of the trace operator. Throughout the paper, we tacitly assume that all the random variables introduced have finite second order moments.

2.2 Notational conventions

Some notational conventions will be followed throughout this paper:

- Matrices and vectors will be written in **boldface**.
- The dimension of each vector and matrix will be indicated at first appearance. For instance, $\mathbf{A}^{(p\times q)}$ denotes a matrix with p rows and q coloumns.
- $\mathbf{A}'$ denotes the transpose of the matrix $\mathbf{A}$.
- $\operatorname{diag}\{\mathbf{a}\}$ denotes a diagonal matrix, the elements of the vector $\mathbf{a}$ forming the diagonal elements.
- $\operatorname{tr}\{\cdot\}$ is the trace operator, that is, $\operatorname{tr}\{\mathbf{A}\}$ is the sum of the main diagonal elements of the square matrix $\mathbf{A}$.
- $(a_k)_{\forall k\in\mathcal{A}}$ denotes the vector $(a_{k_1},\ldots,a_{k_N})'$ where $\mathcal{A}=\{k_1,\ldots,k_N\}$ and $k_1<\cdots<k_N$, and $(a_{kl})_{\forall k\in\mathcal{A}^{(k)}}^{\forall l\in\mathcal{A}^{(l)}}$ denotes the $N\times M$ matrix

$$\begin{pmatrix} a_{k_1l_1} & a_{k_1l_2} & \cdots & a_{k_1l_M} \\ a_{k_2l_1} & a_{k_2l_2} & \cdots & a_{k_2l_M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k_Nl_1} & a_{k_Nl_2} & \cdots & a_{k_Nl_M} \end{pmatrix}$$

  where $\mathcal{A}^{(l)}=\{l_1,\ldots,l_M\}$, $\mathcal{A}^{(k)}=\{k_1,\ldots,k_N\}$, $l_1<\cdots<l_M$ and $k_1<\cdots<k_N$.
- $\mathbf{I}_S$ denotes the $S\times S$ identity matrix.

- We define $\mathrm{Cov}(\mathbf{X}, \mathbf{Y}') = \mathrm{E}(\mathbf{X}\mathbf{Y}') - \mathrm{E}(\mathbf{X})\mathrm{E}(\mathbf{Y}')$. For the sake of simplicity we put $\mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(\mathbf{X}, \mathbf{X}')$. If $\mathbf{X}$ is a scalar we let $\mathrm{Var}(X) = \mathrm{Cov}(\mathbf{X})$.

- δ denotes the Kronecker delta, that is, $\delta_{k,l} = 1$ if $k = l$ and 0 otherwise.

- $I(E)$ is the indicator function of the event E, that is, $I(E) = 1$ if the event E has ocurred and 0 otherwise.

- "a.s." is an abbreviation for "almost surely", or synonymously "with probability 1".

2.3 Invariability of the trace operator

Let $\mathbf{A}^{(r\times s)}$ and $\mathbf{B}^{(s\times r)}$ be two arbitrary matrices with the indicated dimensions. We have

$$\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}, \tag{2.1}$$

that is, the trace operator is invariant under a cyclic permutation of the matrices. Of course, $\mathbf{A}$ and $\mathbf{B}$ may themselves be products of matrices. Remember that, as stated in Section 2.2, the trace of a square matrix is the sum of its main diagonal elements. This implies that the trace of a scalar is the scalar itself and the trace of an identity matrix is its dimension.

Identity (2.1) may easily be verified by straightforward matrix multiplications. A complete argument is given in e.g. Schwartz(1961, pp. 129–130).

2.4 Optimality criterion

Let $\mathbf{m} = (m_1, \ldots, m_p)'$ be an unknown random vector and $\dot{\mathbf{m}} = (\dot{m}_1, \ldots, \dot{m}_p)'$ and $\ddot{\mathbf{m}} = (\ddot{m}_1, \ldots, \ddot{m}_p)'$ two estimators of $\mathbf{m}$. Then we shall say that $\dot{\mathbf{m}}$ is a better estimator of $\mathbf{m}$ than $\ddot{\mathbf{m}}$ if

$$\mathrm{E}(m_i - \dot{m}_i)^2 \leq \mathrm{E}(m_i - \ddot{m}_i)^2 \quad (i = 1, \ldots, p) \tag{2.2}$$

with strict inequality for at least one i. That is, we use elementwise expected quadratic loss as optimality criterion.

2.5 Credibility estimator

Let $\mathbf{X}^{(t\times 1)}$ be an observable random vector and let $\bar{\mathbf{m}}^{(p\times 1)}$ be an estimator of $\mathbf{m}^{(p\times 1)}$ based on $\mathbf{X}$.

Definition 2.1 We shall call $\bar{\mathbf{m}}$ a *linear estimator* of $\mathbf{m}$ based on $\mathbf{X}$ if it can be written in the form

$$\bar{\mathbf{m}} = \mathbf{g} + \mathbf{G}\mathbf{X}$$

where $\mathbf{g}^{(p\times 1)}$ is a non-random vector and $\mathbf{G}^{(p\times t)}$ is a non-random matrix.

Definition 2.2 By a *credibility estimator* we shall mean the best linear estimator with respect to elementwise quadratic loss.

2.6 Recursive credibility estimators

Definition 2.3 For $t = 0, 1, \ldots$ let $\tilde{m}_{t+1}$ be a credibility estimator of the random variable m_{t+1} based on the observations ${}_t\mathbf{X} = (X_1, \ldots, X_t)'$ (${}_0\mathbf{X} = \emptyset$). Then we shall call $\{\tilde{m}_{t+1}\}_{t=0}^{\infty}$ a sequence of *recursive credibility estimators* if its elements can be written in the form

$$\tilde{m}_{t+1} = \alpha_t^{(1)} X_t + \alpha_t^{(2)} \tilde{m}_t + \alpha_t^{(3)}$$

where $\alpha_t^{(i)}$ $(i = 1, 2, 3)$ are non-random scalars for all t.

This way of calculating the credibility estimator will be called *recursive credibility estimation.*

We make use of the concept "recursive credibility estimation" in the same way as in SUNDT(1981).

A more restrictive concept has previously been introduced by GERBER & JONES(1975). They consider what they call "linear credibility formulas of the updating type", which occur by putting $\alpha_t^{(3)} = 0$ and $\alpha_t^{(2)} = 1 - \alpha_t^{(1)}$ for all t in Definition 2.3.

2.7 Empirical recursive credibility estimators

Definition 2.4 By a sequence of *empirical recursive credibility estimators* we shall mean a sequence of recursive credibility estimators in which the structural parameters are replaced by their respective estimators.

It is noteworthy that we do not impose any restrictions on the estimators of the structural parameters in the empirical recursive credibility estimators. This is in contrast with e.g. NORBERG(1980), in which the estimators of the structural parameters in the empirical credibility estimator are assumed to converge in some sense.

2.8 The normal equations

Theorem 2.1 *A linear estimator $\bar{\mathbf{m}}(\mathbf{X})$ of $\mathbf{m}$ based on $\mathbf{X}$ is a credibility estimator if and only if*

$$\mathrm{E}[\bar{\mathbf{m}}(\mathbf{X})] = \mathrm{E}(\mathbf{m}) \tag{2.3}$$

and

$$\mathrm{Cov}(\bar{\mathbf{m}}(\mathbf{X}), \mathbf{X}') = \mathrm{Cov}(\mathbf{m}, \mathbf{X}'). \tag{2.4}$$

If both $\bar{\mathbf{m}}(\mathbf{X})$ and $\bar{\bar{\mathbf{m}}}(\mathbf{X})$ satisfy these equations then

$$\bar{\mathbf{m}}(\mathbf{X}) = \bar{\bar{\mathbf{m}}}(\mathbf{X}) \qquad a.s. \tag{2.5}$$

The credibility estimator $\tilde{\mathbf{m}}(\mathbf{X})$ of $\mathbf{m}$ based on $\mathbf{X}$ satisfies

$$\mathrm{Cov}(\mathbf{m}, \tilde{\mathbf{m}}(\mathbf{X})') = \mathrm{Cov}(\tilde{\mathbf{m}}(\mathbf{X})) = \mathrm{Cov}(\mathbf{m}) - \mathrm{Cov}(\mathbf{m} - \tilde{\mathbf{m}}(\mathbf{X}))$$

Proof. See e.g. Theorem 6.1′ in SUNDT(1991b). □

From Theorem 2.1 we see that the credibility estimator $\tilde{\mathbf{m}}$ of $\mathbf{m}$ based on $\mathbf{X}$ is given by (provided that $\mathrm{Cov}(\mathbf{X})$ is invertible)

$$\tilde{\mathbf{m}} = \mathrm{E}(\mathbf{m}) + \mathrm{Cov}(\mathbf{m}, \mathbf{X}') \left[\mathrm{Cov}(\mathbf{X})\right]^{-1} (\mathbf{X} - \mathrm{E}(\mathbf{X})). \tag{2.6}$$

If the first and second order moments of the distribution of $(\mathbf{m}, \mathbf{X})$ are known, then we can use (2.6) to compute $\tilde{\mathbf{m}}$. However if the dimension of $\mathbf{X}$ is large, the inversion process of $\text{Cov}(\mathbf{X})$ may need a considerable amount of computer resources. In addition, we see from (2.6) that if a new observation is added to $\mathbf{X}$, the whole inversion procedure has to be repeated with our new $\mathbf{X}$.

We will solve this problem by stating models under which the credibility estimators are calculated recursively (Chapters 5–7). This makes it easy to compute the credibility estimator, no matter the dimension of $\mathbf{X}$.

Chapter 3

Linear Sufficiency

3.1 Introduction

The concept of linear sufficiency was first introduced into credibility theory by WITTING(1986,1987). It enables us to answer the question: "Are some of the observations superfluous and can the remaining observations be summarized, without losing any information, when computing the credibility estimator?". An answer to this question is particularly useful in motor insurance where a portfolio typically consists of a large number of car models in which numerous policies are associated with many of these car models.

To make a prospective application of the result in this chapter as easy as possible in other contexts, the notation used in this chapter will differ from the one used in the other chapters of the present paper.

For a discussion on the connection between (ordinary) sufficiency and linear sufficiency, and to get a deeper understanding of linear sufficiency we refer to SUNDT(1991a).

3.2 Sufficient conditions for linear sufficiency

The credibility estimator of the random vector $\mathbf{m}^{(p\times 1)}$ based on the observations $\mathbf{X}^{(t\times 1)}$ is denoted by $\tilde{\mathbf{m}}(\mathbf{X})^{(p\times 1)}$. We then have the following definition, which is identical to the one in WITTING(1986):

Definition 3.1 The linear statistic $\mathbf{T}(\mathbf{X})$ (which is formally a linear transformation $\mathbf{T}$: $\mathbb{R}^n \mapsto \mathbb{R}^r$ with $r < t$) is called *linear sufficient* for $\mathbf{m}$ based on $\mathbf{X}$ if

$$\tilde{\mathbf{m}}(\mathbf{X}) = \tilde{\mathbf{m}}(\mathbf{T}(\mathbf{X})) \qquad \text{a.s.}$$

Suppose we are in a situation where we have data $\mathbf{X}^{(M\times 1)}$ from G independent groups. These groups may e.g. be car models or makes of cars. We are interested in finding the credibility estimator of an unknown random vector $\mathbf{m}^{g,(p\times 1)}$ attached to group g, based on the data vector $\mathbf{X}$. It would be convenient to be able to reduce the dimension of this problem. We therefore divide the data in the following way (see Figure 3.1). We have

$$\mathbf{X} = \left((\mathbf{X}^1)', \ldots, (\mathbf{X}^G)'\right)'$$

where $\mathbf{X}^{g,(m_g\times 1)}$ and $\mathbf{X}^{h,(m_h\times 1)}$ are stochastically independent for $g \neq h$ $(g, h = 1, \ldots, G)$. We write

$$\mathbf{X}^g = \left((\mathbf{X}_1^g)', \ldots, (\mathbf{X}_{r_g}^g)'\right)'$$

Groups	1	$\cdots$	g	$\cdots$	G
Estimands	$\mathbf{m}^1$	$\cdots$	$\mathbf{m}^g$	$\cdots$	$\mathbf{m}^G$
Obser-vations	$\mathbf{X}^1$	$\cdots$	$\left.\begin{array}{cccl} X^g_{r_g 1} & \cdots & X^g_{r_g n^g_{r_g}} & \}\,\mathbf{X}^g_{r_g} \\ X^g_{r_g-1,1} & \cdots & X^g_{r_g-1,n^g_{r_g-1}} & \}\,\mathbf{X}^g_{r_g-1} \\ \vdots & & \vdots & \vdots \\ X^g_{11} & \cdots & X^g_{1n^g_1} & \}\,\mathbf{X}^g_1 \end{array}\right\} \mathbf{X}^g$	$\cdots$	$\mathbf{X}^G$
	$\underbrace{\qquad\qquad\qquad}_{\mathbf{X}}$				

Figure 3.1: How the observations and the unknown random vectors to be estimated are divided into G distinct groups.

where

$$\mathbf{X}_i^{g,(n_i^g\times 1)} = \left(X^g_{i1},\ldots,X^g_{in_i^g}\right)' \quad (i=1,\ldots,r_g) \ ; \ \sum_{i=1}^{r_g} n_i^g = m_g \ ; \ \sum_{g=1}^{G} m_g = M$$

such that

$$\mathbf{T}^g(\mathbf{X}) = \mathbf{T}^g(\mathbf{X}^g) = \left(T_1^g(\mathbf{X}_1^g),\ldots,T_{r_g}^g(\mathbf{X}_{r_g}^g)\right)'$$

where $T_i^g(\mathbf{X}_i^g)$ is a linear transformation $T_i : \mathbb{R}^{n_i^g} \mapsto \mathbb{R}$ with $n_i^g \in \mathbb{N}$ $(i=1,\ldots,r_g \ ; \ g=1,\ldots,G)$.

Theorem 3.1 *Suppose that*

$$\operatorname{Cov}\left(T_j^g(\mathbf{X}_j^g), X_{lk}^g\right) = \operatorname{Cov}\left(T_j^g(\mathbf{X}_j^g), T_l^g(\mathbf{X}_l^g)\right) \tag{3.1}$$

and

$$\operatorname{Cov}(\mathbf{m}^g, X_{lk}^g) = \operatorname{Cov}(\mathbf{m}^g, T_l^g(\mathbf{X}_l^g)) \tag{3.2}$$

for all (j,k,l) and every fixed g. Then $\mathbf{T}^g(\mathbf{X})$ is linear sufficient for $\mathbf{m}^g$ based on $\mathbf{X}$.

Proof. Since $\tilde{\mathbf{m}}^g(\mathbf{T}^g(\mathbf{X}^g))$ is a linear estimator based on $\mathbf{T}^g(\mathbf{X}^g)$, we can write (cf. Definition 2.1)

$$\tilde{\mathbf{m}}^g(\mathbf{T}^g(\mathbf{X}^g)) = \mathbf{b}_0^g + \sum_{j=1}^{r_g} \mathbf{b}_j^g T_j^g(\mathbf{X}_j^g)$$

for non-random vectors $\mathbf{b}_0^{g,(p\times 1)},\ldots,\mathbf{b}_{r_g}^{g,(p\times 1)}$ given by the normal equations (2.3) and (2.4). This gives for fixed $g \in \{1,\ldots,G\}$:

$$\begin{aligned}
\operatorname{Cov}(\tilde{\mathbf{m}}^g(\mathbf{T}^g(\mathbf{X}^g)), X_{lk}^g) &= \operatorname{Cov}(\mathbf{b}_0^g + \sum_{j=1}^{r_g} \mathbf{b}_j^g T_j^g(\mathbf{X}_j^g), X_{lk}^g) = \sum_{j=1}^{r_g} \mathbf{b}_j^g \operatorname{Cov}(T_j^g(\mathbf{X}_j^g), X_{lk}^g) \\
&\overset{(1)}{=} \sum_{j=1}^{r_g} \mathbf{b}_j^g \operatorname{Cov}(T_j^g(\mathbf{X}_j^g), T_l^g(\mathbf{X}_l^g)) \\
&= \operatorname{Cov}(\mathbf{b}_0^g + \sum_{j=1}^{r_g} \mathbf{b}_j^g T_j^g(\mathbf{X}_j^g), T_l^g(\mathbf{X}_l^g)) \\
&= \operatorname{Cov}(\tilde{\mathbf{m}}^g(\mathbf{T}^g(\mathbf{X}^g)), T_l^g(\mathbf{X}_l^g)) \overset{(2)}{=} \operatorname{Cov}(\mathbf{m}^g, T_l^g(\mathbf{X}_l^g)) \\
&\overset{(3)}{=} \operatorname{Cov}(\mathbf{m}^g, X_{lk}^g) \quad \forall (k,l),
\end{aligned}$$

where (1) follows from (3.1), (2) follows from the normal equations (2.4) and (3) follows from (3.2).

For $h \neq g$ we get

$$\mathrm{Cov}(\tilde{\mathbf{m}}^g(\mathbf{T}^g(\mathbf{X}^g)), X_{lk}^h) = \mathbf{0} = \mathrm{Cov}(\mathbf{m}^g, X_{lk}^h) \quad \forall(k,l,h),$$

since group g and h are independent.

The normal equations now give that $\tilde{\mathbf{m}}^g(\mathbf{T}^g(\mathbf{X}^g))$ also is a credibility estimator of $\mathbf{m}^g$ based on $\mathbf{X}$. The identity $\mathbf{T}^g(\mathbf{X}) = \mathbf{T}^g(\mathbf{X}^g)$ and (2.5) gives that $\mathbf{T}^g(\mathbf{X})$ is linear sufficient for $\mathbf{m}^g$ based on $\mathbf{X}$. $\square$

Note that independence between groups is essential to the theorem.

Theorem 3.3 in SUNDT(1991a) is similar to the present theorem. They are not special cases of each other, since they assume a different division of the observations. In addition, SUNDT(1991a) does not make any independence assumptions at this stage.

Theorem 3.1 can be used to prove Lemma 4.1 in SUNDT(1987a).

Chapter 4

Common Model Assumptions and Definitions

4.1 Introduction

Definitions and assumptions made in this chapter will hold for the models in Chapters 5 and 6. With minor changes these will also hold for the model in Chapter 7. These changes will be stated in that chapter.

To simplify notation we assume that we are considering one particular updating. Then some of the quantities defined in Section 4.2 has changed since the last updating. This is the case for e.g. $\mathcal{B}_k$, r_k, n and K. n will increase by 1 from one updating to the next.

4.2 Definitions

n	$=$	number of years the portfolio has been under observation
$\mathcal{A}_j$	$=$	set of indices from all car models still in the portfolio in year j
$\mathcal{A}_{ij}$	$=$	$\mathcal{A}_i \cap \mathcal{A}_j$
$\mathcal{B}_k^{(l)}$	$=$	set of indices from all the years car model k has been in the portfolio up to year l, inclusive
$\mathcal{B}_k$	$=$	$\mathcal{B}_k^{(n)}$
${}^i_jK$	$=$	number of car models in the portfolio in both years i and j (=number of elements in $\mathcal{A}_{ij}$)
K_j	$=$	${}^j_jK$
I_{kj}	$=$	number of policies observed for car model k in year j
$n_k^{(l)}$	$=$	number of years car model k has been in the portfolio up to year l, inclusive (=number of years we have observations from car model k up to year l, inclusive=number of elements in $\mathcal{B}_k^{(l)}$)
n_k	$=$	$n_k^{(n)}$
b_k	$=$	first year we have observations from car model k
r_k	$=$	last year we have observations from car model k
$K^{(l)}$	$=$	number of distinct car models joining the portfolio up to year l, inclusive (=number of elements in the set $\cup_{j=1}^{l}\mathcal{A}_j$)
K	$=$	$K^{(n)}$

I_k	$=$	$\sum_{j \in \mathcal{B}_k} I_{kj}$
$I^{(l)}$	$=$	$\sum_{k=1}^{K^{(l)}} I_k$
I	$=$	$I^{(n)}$
X_{kji}	$=$	total claim amount for policy i of car model k in year j
p_{kji}	$=$	earned premium for policy i of car model k in year j
f_{kj}	$=$	old factor for car model k in year j
$w_{kji} = p_{kji}/f_{kj}$	$=$	our measure of the risk volume for policy i of car model k in year j
$Y_{kji} = X_{kji}/w_{kji}$	$=$	claim amount per unit of risk volume for policy i of car model k in year j
Θ_{kj}	$=$	unobservable random parameter characterizing car model k in year j
$\Theta_k = \{\Theta_{kj}\}_{j=b_k}^{\infty}$	$=$	sequence of unobservable random parameters characterizing car model k.

Further

$$
\begin{array}{llllll}
w_{kj\cdot} & = & \sum_{i=1}^{I_{kj}} w_{kji} & \mathbf{Y}_j^{((\sum_{k\in\mathcal{A}_j} I_{kj})\times 1)} & = & \left(\mathbf{Y}'_{kj}\right)_{\forall k\in\mathcal{A}_j} \\
w_{\cdot j\cdot} & = & \sum_{k\in\mathcal{A}_j} w_{kj\cdot} & {}_l\mathbf{Y}^{(I^{(l)}\times 1)} & = & (\mathbf{Y}'_1,\ldots,\mathbf{Y}'_l)' \\
Y_{kj\cdot} & = & w_{kj\cdot}^{-1}\sum_{i=1}^{I_{kj}} w_{kji}Y_{kji} & {}_l\mathbf{Y}_{k\Sigma}^{(n_k^{(l)}\times 1)} & = & (Y_{kj\cdot})_{\forall j\in\mathcal{B}_k^{(l)}} \\
\mathbf{Y}_{kj}^{(I_{kj}\times 1)} & = & (Y_{kj1},\ldots,Y_{kjI_{kj}})' & \mathbf{D}_j^{(K_j\times K_j)} & = & \operatorname{diag}\left\{(w_{kj\cdot}/w_{\cdot j\cdot})_{\forall k\in\mathcal{A}_j}\right\}
\end{array}
$$

and for $m \geq l$ $(i = l, m)$

$$
\begin{array}{llllll}
{}_l^m\mathbf{Y}_{\Sigma i}^{({}_l^m K\times 1)} & = & (Y_{ki\cdot})_{\forall k\in\mathcal{A}_{lm}} & {}_l^m\mathbf{T}_i^{({}_l^m K\times q_i)} & = & (\mathbf{t}_{ki})_{\forall k\in\mathcal{A}_{lm}} \\
\mathbf{Y}_{\Sigma j}^{(K_j\times 1)} & = & {}_j^j\mathbf{Y}_{\Sigma j} & \mathbf{T}_j^{(K_j\times q_j)} & = & {}_j^j\mathbf{T}_j.
\end{array}
$$

For an explanation of $\mathbf{t}_{kj}$ we refer to Section 5.2.

Our choice of measure of risk volume may be criticized. During this study the other rating factors are assumed to be fixed. But a higher value of the other rating factors will give a higher value of our risk volume. In particular, the measure of risk volume depends on inflation. This is clearly unfortunate since the measure of risk volume may increase even though the exposure volume has not. An advantage of our measure of risk volume is that by dividing X_{kji} by w_{kji}, Y_{kji} becomes (to a great extent) independent of inflation. Further, $w_{kj\cdot}$ is increasing in I_{kj}, and $w_{kj\cdot}$ depends on for how long car model k has been in the portfolio in year j (since p_{kji} is the *earned* premium), which are reasonable. We are therefore willing to use w_{kji} as our measure of risk volume.

4.3 Assumptions

4.1. $X_{1j_1i_1}, \ldots, X_{Kj_Ki_K}$ are stochastically independent for all (j_k, i_k) $(j_k \in \mathcal{B}_k\ ;\ i_k = 1, \ldots, I_{kj_k})$.

4.2. X_{kji} and $X_{kj'i'}$ are conditionally independent given Θ_k for $(i, j') \neq (i', j)$ $(i = 1, \ldots, I_{kj}\ ;\ i' = 1, \ldots, I_{kj'}\ ;\ j, j' \in \mathcal{B}_k\ ;\ k = 1, \ldots, K)$.

4.3. The X_{kji}'s depend on Θ_k only through Θ_{kj}.

4.4. $\Theta_1, \ldots, \Theta_K$ are independent.

The consequence of both the first and last assumption is that one can not learn anything about the claims of one car model by observing the claims of another car model in the portfolio. The second assumption indicates that the claims from one and the same car model depend on each other only because they are influenced by the same unobservable risk characteristics Θ_k. The third assumption says that all of the unobservable risk characteristics for car model k in year j are contained in Θ_{kj}.

Chapter 5

Time-heterogeneous Model

5.1 Introduction

In this chapter we will present a model which allows the structural parameters to vary over time. The present author feels that this is a reasonable approach since the processes which develop the claims in motor insurance are likely to evolve as time passes.

5.2 Assumptions

The following assumptions will be specific for the present model:

5.1. $\mathrm{E}(Y_{kji}|\Theta_{kj}) = m_{kj}(\Theta_{kj})$

5.2. $\mu_{kj} = \mathrm{E}[m_{kj}(\Theta_{kj})] = \mathbf{t}'_{kj}\boldsymbol{\beta}_j$

5.3. $\mathrm{Var}(Y_{kji}|\Theta_{kj}) = s_j^2(\Theta_{kj})/v_{kji}$ where, for the time being, $v_{kji} = w_{kji}$

5.4. $\mathrm{E}[s_j^2(\Theta_{kj})] = \varphi_j$

5.5. $\mathrm{Cov}(m_{ki}(\Theta_{ki}), m_{kj}(\Theta_{kj})) = \lambda_{ij}$, $\lambda_j = \lambda_{jj}$

5.6. $\lambda_{i+1,j} = \varrho_i \lambda_{ij}$ $(i \geq j)$

5.7. $\mathrm{Corr}(m_{ki}(\Theta_{ki}), m_{kj}(\Theta_{kj}))) = \pi_{ij} \in [0,1]$

We will call $m_{kj}(\Theta_{kj})$ the *risk level* of car model k in year j.

$\mathbf{t}_{kj}^{(q_j \times 1)}$ is a known vector of the technical variables (including a constant term) of car model k in year j. In Storebrand we used $\mathbf{t}_{kj} = (1, \text{engine power}, \text{price/weight})'$ in vehicle damage and $\mathbf{t}_{kj} = (1, \text{engine power})'$ in liability insurance. The dependence of j in $\mathbf{t}_{kj}$ indicates that the technical variables of the car model may change from year to year. The price is likely to change as time passes, but other technical variables may also change. This depends heavily on the definition of a car model. If a car model has changed considerably since it was introduced to the market, one might question whether it still is the same car model or not. The dependence of j in q_j indicates that also the number of technical variables is allowed to vary over time. One could face the possibility that after a number of years one finds another set of technical variables which describes the expected risk level better than the "old" set, in some sense. We shall see that this is exactly the case in Chapter 7.

$\beta_j^{(q_j \times 1)}$ is an unknown regressor vector.
The reason for introducing v_{kji} will become clear in Section 11.2.
From (A.7) and Assumption 5.1 we can write

$$\mathrm{Var}(Y_{kji}) = \mathrm{E}[\mathrm{Var}(Y_{kji}|\Theta_{kj})] + \mathrm{Var}(m_{kj}(\Theta_{kj})).$$

Hence, the total variation of the Y_{kji}'s can be split up in two parts; the first part describes the pure random fluctuation of the Y_{kji}'s and the second part describes the variation of the risk levels.

It is Assumption 5.6 that makes it possible for us to derive a recursive procedure for the calculation of the credibility estimators. (A.13) gives ($i \geq j$):

$$\pi_{i+1,j} = \pi_{i+1,i}\pi_{ij}. \tag{5.1}$$

By comparing (5.1) with Assumption 5.6 we see that in the present model the correlation structure is the same as the covariance structure. Since (5.1) and Assumption 5.6 are equivalent expressions, (5.1) is also a sufficient condition for the credibility estimators to be recursive (Definition 2.3 explains what we mean by recursive credibility estimators). For $i > j$, (5.1) is both a sufficient and necessary condition (cf. SUNDT(1981, pp. 5–6)). By repeated use of (5.1), we get

$$\pi_{i+1,j} = \prod_{k=j}^{i} \pi_{k+1,k}. \quad (i \geq j) \tag{5.2}$$

This means that we find the correlation between the risk levels in two distinct years by multiplying the in between correlations in risk levels from one year to the next.

Let us assume that $m < l \leq i$. Then from (5.2) we have

$$\pi_{i+1,m} = \left(\prod_{k=l}^{i} \pi_{k+1,k}\right) \prod_{k=m}^{l-1} \pi_{k+1,k} = \pi_{i+1,l} \prod_{k=m}^{l-1} \pi_{k+1,k}. \tag{5.3}$$

From Assumption 5.7 we have that $0 \leq \pi_{ij} \leq 1 \ \forall (i,j)$, implying that $0 \leq \prod_{k=m}^{l-1} \pi_{k+1,k} \leq 1$. (5.3) gives

$$\pi_{i+1,m} \leq \pi_{i+1,l}. \quad (m < l \leq i) \tag{5.4}$$

This means that the correlation between the risk levels in two distinct years increases when the distance between the years decreases (in the weak sense).

The reason for restricting π_{ij} to the interval $[0,1]$ in Assumption 5.7 should be obvious. The present author feels that it is unreasonable to allow the risk levels to be negatively correlated.

5.3 Linear sufficient statistic

Assume that we, at the end of year n, want to estimate the risk levels $(m_{k,n+1}(\Theta_{k,n+1}))_{\forall k \in \mathcal{A}_{n+1}}$ using all the available observations ${}_n\mathbf{Y}$ up to that point in time. From (2.6) we see that the credibility estimator of the vector $(m_{k,n+1}(\Theta_{k,n+1}))_{\forall k \in \mathcal{A}_{n+1}}$ based on ${}_n\mathbf{Y}$ is the same as the vector of the credibility estimators of $m_{k,n+1}(\Theta_{k,n+1})$ based on ${}_n\mathbf{Y}$ for all $k \in \mathcal{A}_{n+1}$. We will therefore restrict our attention to finding the credibility estimator of $m_{s,n+1}(\Theta_{s,n+1})$ based on ${}_n\mathbf{Y}$, with $s \in \mathcal{A}_{n+1}$ fixed.

In practice the vector ${}_n\mathbf{Y}$ will be extremely large, even for small n. For computational reasons it would be very helpful if we could reduce the dimension of the estimation problem described above. We shall see that we are in fact able to do this, as a consequence of our model specifications.

Theorem 5.1 *The statistic* ${}_t\mathbf{Y}_{s\Sigma}$ *is linear sufficient for* $m_{su}(\Theta_{su})$ *based on* ${}_t\mathbf{Y}$ $(t, u \in \{b_s, b_s+1, \ldots\})$.

Proof. Let $t, u \in \{b_s, b_s+1, \ldots\}$ be fixed. Since the division of the observations is identical to the one in Section 3.2, it is sufficient to show that the conditions of Theorem 3.1 are satisfied in the present model. We have from (A.8)

$$\mathrm{Cov}(Y_{sj\cdot}, Y_{sli}) = \delta_{j,l}\varphi_j/v_{sj\cdot} + \lambda_{jl} \quad (j, l \in \mathcal{B}_s^{(t)}\ ;\ i = 1, \ldots, I_{sl})$$

and from (A.9)

$$\mathrm{Cov}(Y_{sj\cdot}, Y_{sl\cdot}) = \delta_{j,l}\varphi_j/v_{sj\cdot} + \lambda_{jl}. \quad (j, l \in \mathcal{B}_s^{(t)})$$

So (3.1) is fulfilled. Further, from (A.10) and (A.11), respectively, we have

$$\begin{aligned} \mathrm{Cov}(m_{s,u}(\Theta_{s,u}), Y_{sli}) &= \lambda_{u,l} \quad (l \in \mathcal{B}_s^{(t)}\ ;\ i = 1, \ldots, I_{sl}) \\ \mathrm{Cov}(m_{s,u}(\Theta_{s,u}), Y_{sl\cdot}) &= \lambda_{u,l} \quad (l \in \mathcal{B}_s^{(t)}) \end{aligned}$$

and so (3.2) is also satisfied. Since we have assumed independence between car models, the theorem follows from Theorem 3.1. □

Theorem 5.1 allows us to base the credibility estimator of $m_{s,u}(\Theta_{s,u})$ on the summarized observations ${}_t\mathbf{Y}_{s\Sigma}$ $(t, u \in \{b_s, b_s+1, \ldots\})$

5.4 Recursive credibility estimators

Assume that we are at the end of year t. In order to assess the premium for car model s in year $t+1$, we want to estimate $m_{s,t+1}(\Theta_{s,t+1})$ based on the observations ${}_t\mathbf{Y}_{s\Sigma}$. We will do this by deriving the credibility estimator of the risk level. As we will see, this estimator is easy to compute in our present model.

Let $\tilde{m}_{s,i|j}$ be the credibility estimator of $m_{si}(\Theta_{si})$ based on ${}_j\mathbf{Y}_{s\Sigma}$, and define $\psi_{s,i|j} = \mathrm{E}[(m_{si}(\Theta_{si}) - \tilde{m}_{s,i|j})^2]$. (We put ${}_{(b_s-1)}\mathbf{Y}_{s\Sigma} = \emptyset$).

Theorem 5.2 *We have* $(t = b_s, b_s+1, \ldots)$

$$\tilde{m}_{s,b_s|b_s-1} = \mu_{sb_s} \qquad \psi_{s,b_s|b_s-1} = \lambda_{b_s}$$

$$\psi_{s,t+1|t} = \varrho_t^2\left[\frac{\psi_{s,t|t-1}\,\varphi_t}{v_{st\cdot}\psi_{s,t|t-1} + \varphi_t} - \lambda_t\right] + \lambda_{t+1} \tag{5.5}$$

$$\tilde{m}_{s,t+1|t} = \varrho_t\left[\frac{v_{st\cdot}\psi_{s,t|t-1}}{v_{st\cdot}\psi_{s,t|t-1} + \varphi_t}Y_{st\cdot} + \frac{\varphi_t}{v_{st\cdot}\psi_{s,t|t-1} + \varphi_t}\tilde{m}_{s,t|t-1} - \mu_{st}\right] + \mu_{s,t+1}. \tag{5.6}$$

Proof. Since s is fixed, the theorem follows from (11) in SUNDT(1981). □

Since μ_{sl} is the credibility estimator of $m_{sl}(\Theta_{sl})$ based on no observations, we will call μ_{sl} the prior estimator in year l $(l = b_s, b_s+1, \ldots)$.

An interpretation of (5.5) and (5.6) by the aid of (5.10) and (5.11), is given below.

Using Definition 2.3 we see that the credibility estimators in Theorem 5.2 are recursive credibility estimators.

Theorem 5.2 presents estimators of the risk levels of the car model in a future time period. Therefore $\tilde{m}_{s,t|t-1}$ $(t = b_s, b_s + 1, \ldots)$ are actually predictors.

From (5.6) we see that $\tilde{m}_{s,t+1|t}$ is a weighted sum of $Y_{st\cdot}$, $\tilde{m}_{s,t|t-1}$, μ_{st} and $\mu_{s,t+1}$. That is, a weighted sum of the new observation, the old credibility estimator (which contains the past experience about the car model), the prior estimator from the last year, and the prior estimator of the next year for the car model.

In order to interpret the expressions in Theorem 5.2 more easily we state the following corollary. The proof is an easy consequence of the normal equations and the fact that $\tilde{m}_{s,t|t-1}$ is a credibility estimator which was shown in Theorem 5.2. Corollary 5.1 will therefore not be proven here. An argument is given in Section 8.3 by use of the Kalman filter.

Corollary 5.1 *We have* $(t = b_s, b_s + 1, \ldots)$

$$\begin{aligned} \psi_{s,t|t} &= (1 - \zeta_{st})\psi_{s,t|t-1} & (5.7) \\ \tilde{m}_{s,t|t} &= \zeta_{st} Y_{st\cdot} + (1 - \zeta_{st})\tilde{m}_{s,t|t-1} & (5.8) \end{aligned}$$

where

$$\zeta_{st} = \frac{\psi_{s,t|t-1}}{\psi_{s,t|t-1} + \varphi_t / v_{st\cdot}}. \tag{5.9}$$

From (5.8) we see that $\tilde{m}_{s,t|t}$ is a weighted mean of the observation from year t and the credibility estimator of $m_{st}(\Theta_{st})$ based on observations up to time $t-1$. The numerator on the right hand side of (5.9) is simply the estimation error of $\tilde{m}_{s,t|t-1}$. We shall take a closer look at the denominator. We first consider

$$\begin{aligned} \mathrm{E}(Y_{st\cdot} - m_{st}(\Theta_{st}))^2 &\overset{(1)}{=} \mathrm{E}\{\mathrm{E}[(Y_{st\cdot} - m_{st}(\Theta_{st}))^2 | \Theta_{st}]\} \\ &\overset{(2)}{=} \mathrm{E}[\mathrm{Var}(Y_{st\cdot} | \Theta_{st})] \overset{(3)}{=} \frac{\varphi_t}{v_{st\cdot}} \end{aligned}$$

where we in (1) have used (A.15), in (2) we have used the definition of conditional variance, and in (3) we have used (A.6) together with Assumption 5.4. Therefore the denominator on the right hand side of (5.9) is the sum of the estimation error of $Y_{st\cdot}$ (considered as an estimator of $m_{st}(\Theta_{st})$) and of the estimation error of $\tilde{m}_{s,t|t-1}$. So the weight ζ_{st} given to $Y_{st\cdot}$ in (5.8) is the fraction between the estimation error of $\tilde{m}_{s,t|t-1}$ and the sum of the estimation errors of both $Y_{st\cdot}$ and $\tilde{m}_{s,t|t-1}$. On the other hand since $1-\zeta_{st} = (\varphi_t/v_{st\cdot})/(\psi_{s,t|t-1}+\varphi_t/v_{st\cdot})$, the weight given to $\tilde{m}_{s,t|t-1}$ is the fraction between the estimation error of $Y_{st\cdot}$ and the sum of the estimation errors of both $Y_{st\cdot}$ and $\tilde{m}_{s,t|t-1}$. This is very reasonable. Since $\tilde{m}_{s,t|t}$ is a linear estimator of $m_{st}(\Theta_{st})$ based on observations up to year t, it is sensible that $\tilde{m}_{s,t|t}$ is a weighted sum of an estimator of $m_{st}(\Theta_{st})$ based on observations from year t ($Y_{st\cdot}$) and an estimator of $m_{st}(\Theta_{st})$ based on observations up to year $t-1$ ($\tilde{m}_{s,t|t-1}$). The weight given to $Y_{st\cdot}$ increases when its estimation error decreases (that is, $Y_{st\cdot}$ becomes a better estimator) and the weight given to $\tilde{m}_{s,t|t-1}$ increases when its estimation error decreases ($\tilde{m}_{s,t|t-1}$ becomes a better estimator).

(5.7) implies that $\psi_{s,t|t} \leq \psi_{s,t|t-1}$ (with equality if and only if $v_{st\cdot} = 0$ or $\psi_{s,t|t-1} = 0$; in the former case there are no observations from year t and in the latter $\tilde{m}_{s,t|t-1}$ is a "perfect" estimator and hence $Y_{st.}$ does not give any further information) which is reasonable since $\tilde{m}_{s,t|t}$ is based on more information than $\tilde{m}_{s,t|t-1}$ and they estimate the same quantity.

Using Corollary 5.1 we can rewrite (5.5) and (5.6) as

$$\psi_{s,t+1|t} = \varrho_t^2 \left(\psi_{s,t|t} - \lambda_t\right) + \lambda_{t+1} \tag{5.10}$$

$$\tilde{m}_{s,t+1|t} = \varrho_t \left(\tilde{m}_{s,t|t} - \mu_{st}\right) + \mu_{s,t+1}, \tag{5.11}$$

respectively. From (5.11) we see that $\tilde{m}_{s,t+1|t}$ equals the sum of the prior estimator in year $t+1$ and a correction term. This correction term is the difference between the credibility estimator of $m_{st}(\Theta_{st})$ and the prior estimator in year t of the same quantity multiplied by a factor which is proportional to the correlation between $m_{st}(\Theta_{st})$ and $m_{s,t+1}(\Theta_{s,t+1})$. This is also reasonable. An analogous argument can be applied to $\psi_{s,t+1|t}$.

We have from (5.10) that $\psi_{s,t+1|t}$ not necessarily will be less than $\psi_{s,t|t}$. This depends on the values of ϱ_t, λ_t and λ_{t+1} in combination with the value of $\psi_{s,t|t}$. This may seem surprising at first glance since according to (5.11) $\tilde{m}_{s,t+1|t}$ is apparently based on more information than $\tilde{m}_{s,t|t}$. However, this is only apparent. The point is that $\tilde{m}_{s,t+1|t}$ and $\tilde{m}_{s,t|t}$ do not estimate the same quantity so that if $\mu_{s,t+1}$ is not a better estimator of $m_{s,t+1}(\Theta_{s,t+1})$ than μ_{st} is of $m_{st}(\Theta_{st})$, then $\psi_{s,t+1|t}$ may exceed $\psi_{s,t|t}$.

Chapter 6

Time-homogeneous Model

6.1 Introduction

An interesting special case of the model in Chapter 5 is the time-homogeneous model. Here we assume the structural parameters to be constant over time. The present author feels that this is a more unrealistic assumption in motor insurance, than the one in Chapter 5. After all we live in a changing world. This is particularly true for the world of car models. But the present model possesses some interesting properties which are worth investigating more closely.

Similar properties can be derived within the time-heterogeneous model as well, but only under restrictive assumptions which are difficult to interpret.

Since the present model is a special case of the model in Chapter 5, all the results of Chapter 5 will still hold. In particular, Theorem 5.1 will hold in our present model.

6.2 Definitions and assumptions

These definitions are made in the present model:

$$
\begin{array}{lcllcl}
N & = & \sum_{k=1}^{K} n_k & \mathbf{T}^{(k),(n_k\times q)} & = & (\mathbf{t}_k,\ldots,\mathbf{t}_k)' \\
w_{k\cdot\cdot} & = & \sum_{j\in\mathcal{B}_k} w_{kji} & \mathbf{T}^{(N\times q)} & = & \left((\mathbf{T}^{(1)})',\ldots,(\mathbf{T}^{(K)})'\right)' \\
w_{\cdot\cdot\cdot} & = & \sum_{k=1}^{K} w_{k\cdot\cdot} & \mathbf{Y}_{\Sigma}^{(N\times 1)} & = & \left({}_{r_1}\mathbf{Y}'_{1\Sigma},\ldots,{}_{r_K}\mathbf{Y}'_{K\Sigma}\right)' \\
\mathbf{D}^{(k),(n_k\times n_k)} & = & \operatorname{diag}\left\{(w_{kj\cdot}/w_{\cdot\cdot\cdot})_{\forall j\in\mathcal{B}_k}\right\} & \mathbf{D}^{(N\times N)} & = & \operatorname{diag}\left\{\mathbf{D}^{(1)},\ldots,\mathbf{D}^{(K)}\right\}
\end{array}
$$

We state the following assumptions for the present model:

6.1. $\mathrm{E}(Y_{kji}|\Theta_{kj}) = m_k(\Theta_{kj})$

6.2. $\mu_k = \mathrm{E}[m_k(\Theta_{kj})] = \mathbf{t}_k'\boldsymbol{\beta}$

6.3. $\mathrm{Var}(Y_{kji}|\Theta_{kj}) = s^2(\Theta_{kj})/v_{kji}$ where $v_{kji} = w_{kji}$

6.4. $\mathrm{E}[s^2(\Theta_{kj})] = \varphi$

6.5. $\mathrm{Cov}(m_k(\Theta_{ki}), m_k(\Theta_{kj})) = \varrho^{|i-j|}\lambda$

6.6. $\mathrm{Corr}(m_k(\Theta_{ki}), m_k(\Theta_{kj}))) = \pi_{ij} \in [0,1]$

Assumption 6.1 indicates that the risk level varies from year to year only through the development of the latent risk parameter.

Of course, in this model the dimension q of $\mathbf{t}_k$ is independent of the year. $\beta^{(q\times 1)}$ is an unknown regressor vector.

Further, in this model we have $\pi_{ij} = \varrho^{|i-j|}$, which depends on i and j only through $|i-j|$. This implies that $\pi_{j+1,j} = \varrho$, independent of j, which again implies that $0 \le \varrho \le 1$. So π_{ij} is non-increasing in $|i-j|$.

6.3 Recursive credibility estimators

Let $\tilde{m}_{s,t+1|t}$ be the credibility estimator of $m_s(\Theta_{s,t+1})$ based on ${}_t\mathbf{Y}_{s\Sigma}$, and define $\psi_{s,t+1|t} = \mathrm{E}(m_s(\Theta_{s,t+1}) - \tilde{m}_{s,t+1|t})^2$.

In this model Theorem 5.2 becomes:

Corollary 6.1 *We have* $(t = b_s, b_s+1, \ldots)$

$$\tilde{m}_{s,b_s|b_s-1} = \mu_s \qquad \psi_{s,b_s|b_s-1} = \lambda \tag{6.1}$$

$$\psi_{s,t+1|t} = \varrho^2 \frac{\psi_{s,t|t-1}\varphi}{v_{st\cdot}\psi_{s,t|t-1} + \varphi} + \left(1-\varrho^2\right)\lambda \tag{6.2}$$

$$\tilde{m}_{s,t+1|t} = \varrho\left[\frac{v_{st\cdot}\psi_{s,t|t-1}}{v_{st\cdot}\psi_{s,t|t-1}+\varphi}Y_{st\cdot} + \frac{\varphi}{v_{st\cdot}\psi_{s,t|t-1}+\varphi}\tilde{m}_{s,t|t-1}\right] + (1-\varrho)\mu_s. \tag{6.3}$$

Proof. Substitute Assumptions 6.1–6.5 into Theorem 5.2. □

From (6.2) and (6.3) we see that for $\varrho = 0$ we have $\psi_{s,t+1|t} = \lambda$ and $\tilde{m}_{s,t+1|t} = \mu_s$. Since $\mathrm{Cov}(m_s(\Theta_{s,t+1}), Y_{sj\cdot}) = \varrho^{t+1-j}\lambda$ $(j \in \mathcal{B}_s^{(t)})$, we have in this case that $m_s(\Theta_{s,t+1})$ is uncorrelated with all previous observations. This means that there is no *linear* relationship between $m_s(\Theta_{s,t+1})$ and previous observations. Since we use a linear estimator, the previous observations shall not be included in the estimator of $m_s(\Theta_{s,t+1})$. Technically, we are in the same position when we have no data from the car model. This motivates the initial values (6.1) of the recursions.

From (6.3) we have that $\tilde{m}_{s,t+1|t}$ is a weighted mean of $Y_{st\cdot}$, $\tilde{m}_{s,t|t-1}$ and μ_s, that is, the new observation, the estimator from the last updating and prior information. The weight given to the new observation is increasing in the estimation error of $\tilde{m}_{s,t|t-1}$, the risk volume and the correlation between $m_s(\Theta_{st})$ and $m_s(\Theta_{s,t+1})$. This is reasonable since if $\tilde{m}_{s,t|t-1}$ is a good estimator then the new observation $Y_{st\cdot}$ can give little additional information about $m_s(\Theta_{st})$ and the weight given to $Y_{st\cdot}$ should be small. Moreover, if the risk volume in year t is large then $Y_{st\cdot}$ contains a lot of information and large weight should be given to it. Since $Y_{st\cdot}$ is an unbiased estimator of $m_s(\Theta_{st})$ (and $\tilde{m}_{s,t+1|t}$ is an unbiased estimator of $m_s(\Theta_{s,t+1})$), larger weight should be given to $Y_{st\cdot}$ the larger the correlation between $m_s(\Theta_{st})$ and $m_s(\Theta_{s,t+1})$ is. This last argument applies to $\tilde{m}_{s,t|t-1}$ as well. In addition, the weight given to $\tilde{m}_{s,t|t-1}$ is decreasing in the risk volume and its estimation error. Less weight is given to $Y_{st\cdot}$ the larger the pure random fluctuation $\varphi/v_{st\cdot}$ of $Y_{st\cdot}$ is and consequently more weight is given to $\tilde{m}_{s,t|t-1}$. This is clearly reasonable. Finally, the weight given to the prior information μ_s is decreasing in the correlation between $m_s(\Theta_{st})$ and $m_s(\Theta_{s,t+1})$. This is rational since if the correlation is large, then $Y_{st\cdot}$ and $\tilde{m}_{s,t|t-1}$ contains much information about $m_s(\Theta_{s,t+1})$. In addition $\tilde{m}_{s,t|t-1}$ already contains prior information so if ϱ is close to 1 prior information added to $Y_{st\cdot}$ and $\tilde{m}_{s,t|t-1}$ should be given little weight.

Similar arguments can be given to justify the expression (6.2) for $\psi_{s,t+1|t}$.

We see from (6.3) that observations from different years have different weights. In Theorem 6.1 we shall see that under an intuitively reasonable assumption, our recursive credibility estimators give less weight to older observations than to new ones.

To keep things simple we will for the rest of this chapter assume that we have observations from car model s for all the years $b_s, \ldots, t$; that is, $v_{sj\cdot} > 0$ $(j = b_s, \ldots, t)$.

Let the coefficients $\alpha_{t0}, \alpha_{tb_s}, \ldots, \alpha_{tt}$ be defined by

$$\tilde{m}_{s,t+1|t} = \alpha_{t0} + \sum_{j=b_s}^{t} \alpha_{tj} Y_{sj\cdot}.$$

We then have the following:

Theorem 6.1 *Suppose that for* $j \in \{b_s, \ldots, t-1\}$

$$v_{sj\cdot} \leq v_{s,j+1,\cdot} \tag{6.4}$$

$$0 < \varrho < 1, \;\; \lambda > 0 \;\; \textit{and} \;\; \varphi > 0. \tag{6.5}$$

Then

$$0 < \alpha_{tj} < \alpha_{t,j+1} < 1.$$

Proof. The proof is parallel to the one of Theorem 6.4 in SUNDT(1991b). From (6.3) we see that for $j = b_s - 1, \ldots, t-2$ we have

$$\alpha_{tt} = \varrho \frac{v_{st\cdot}\psi_{s,t|t-1}}{v_{st\cdot}\psi_{s,t|t-1} + \varphi} \tag{6.6}$$

$$\alpha_{t,j+1} = \varrho \frac{\varphi}{v_{st\cdot}\psi_{s,t|t-1} + \varphi} \cdots \varrho \frac{\varphi}{v_{s,j+2\cdot}\psi_{s,j+2|j+1} + \varphi} \varrho \frac{v_{s,j+1\cdot}\psi_{s,j+1|j}}{v_{s,j+1\cdot}\psi_{s,j+1|j} + \varphi}. \tag{6.7}$$

These equations give that

$$\alpha_{tj} = \gamma_j \alpha_{t,j+1} \quad (j \in \{b_s, \ldots, t-1\}) \tag{6.8}$$

with

$$\gamma_j = \frac{\varrho \frac{v_{sj\cdot}\psi_{s,j|j-1}}{v_{sj\cdot}\psi_{s,j|j-1} + \varphi} \varrho \frac{\varphi}{v_{s,j+1,\cdot}\psi_{s,j+1|j} + \varphi}}{\varrho \frac{v_{s,j+1,\cdot}\psi_{s,j+1|j}}{v_{s,j+1,\cdot}\psi_{s,j+1|j} + \varphi}}.$$

That is,

$$\gamma_j = \frac{\varrho v_{sj\cdot}\psi_{s,j|j-1}\varphi}{v_{s,j+1,\cdot}\psi_{s,j+1|j}\left(v_{sj\cdot}\psi_{s,j|j-1} + \varphi\right)}$$

and substitution of (6.2) gives

$$\gamma_j = \frac{\varrho v_{sj\cdot}\psi_{s,j|j-1}\varphi}{v_{s,j+1,\cdot}\left[\varrho^2\psi_{s,j|j-1}\varphi + \left(1-\varrho^2\right)\lambda\left(v_{sj\cdot}\psi_{s,j|j-1} + \varphi\right)\right]}. \tag{6.9}$$

From (6.2) we see that

$$\psi_{s,j|j-1} > \left(1 - \varrho^2\right)\lambda > 0$$

where the last inequality is a result of (6.5). We are therefore allowed to divide both numerator and denominator in (6.9) by $\psi_{s,j|j-1}$, and get

$$\gamma_j = \frac{\varrho v_{sj\cdot}\varphi}{v_{s,j+1,\cdot}\left[\varrho^2\varphi + \left(1-\varrho^2\right)\lambda\left(v_{sj\cdot} + \frac{\varphi}{\psi_{s,j|j-1}}\right)\right]}. \tag{6.10}$$

Since the credibility estimator $\tilde{m}_{s,j|j-1}$ is based on at least as much information as μ_s, $\tilde{m}_{s,j|j-1}$ is at least just as good an estimator of $m_s(\Theta_{sj})$ as μ_s is. Thus,

$$\psi_{s,j|j-1} = \mathrm{E}\left(m_s(\Theta_{sj}) - \tilde{m}_{s,j|j-1}\right)^2 \leq \mathrm{E}\left(m_s(\Theta_{sj}) - \mu_s\right)^2 = \lambda,$$

and

$$\begin{aligned}\gamma_j &\leq \frac{\varrho v_{sj\cdot}\varphi}{v_{s,j+1,\cdot}\left[\varrho^2\varphi + (1-\varrho^2)\lambda\left(v_{sj\cdot} + \frac{\varphi}{\lambda}\right)\right]} \\ &= \frac{v_{sj\cdot}}{v_{s,j+1,\cdot}}\frac{\varrho\varphi}{\left[\varphi + (1-\varrho^2)\lambda v_{sj\cdot}\right]} \overset{(1)}{<} \frac{v_{sj\cdot}}{v_{s,j+1,\cdot}}1 \overset{(2)}{\leq} 1, \quad (j \in \{b_s, \ldots, t-1\})\end{aligned}$$

where (1) follows from (6.5), and (2) from (6.4). On the other hand we see from (6.10) that $\gamma_j > 0$ and therefore $0 < \gamma_j < 1$. (6.6) and (6.7) combined with (6.4) and (6.5) give

$$0 < \alpha_{t,j+1} < 1.$$

The result follows from (6.8). □

Essentially Theorem 6.1 says that if the risk volume is non-decreasing from one year to another, then the weight given to the observations from these years, in the same credibility estimator, are strictly increasing. This is typically the case for e.g. car models recently introduced to the market.

Repeated use of Theorem 6.1 gives that if $v_{sb_s\cdot} \leq v_{s,b_s+1,\cdot} \leq \cdots \leq v_{st\cdot}$, $0 < \varrho < 1$, $\lambda > 0$ and $\varphi > 0$ then $0 < \alpha_{tb_s} < \alpha_{t,b_s+1} < \cdots < \alpha_{tt} < 1$. This assumption is very strong, but it illustrates in which way the risk volumes influence on the weights given to the observations.

Another property of the recursive credibility estimators is that the weight given to the observation from one particular year is non-increasing in the year of the risk level to be estimated:

Theorem 6.2 *We have for $j \in \{b_s, \ldots, t-1\}$*

$$\alpha_{tj} \leq \alpha_{t-1,j}.$$

Proof. From (6.3) we have $\alpha_{tj} = \varrho\frac{\varphi}{v_{st\cdot}\psi_{s,t|t-1}+\varphi}\alpha_{t-1,j} \leq \alpha_{t-1,j}$. □

It would be interesting to find conditions under which the credibility estimator is a better predictor as time passes (meaning, the more observations we get). We shall see in the subsequent that in this respect it is the development of the risk volumes that is important.

For the rest of this chapter we assume $\varphi > 0$. (This gives no loss of generality since $\varphi = 0$ is a degenerate case).

Theorem 6.3 *Let $l \in \{b_s+1, b_s+2, \ldots\}$ be fixed. The sequence $\{\psi_{s,j|j-1}\}_{j=l-1}^{\infty}$ is non-increasing if*

$$v_{s,l-1,\cdot} \leq v_{sl\cdot} \leq \cdots \tag{6.11}$$

and

$$\psi_{s,l-1|l-2} \geq \psi_{s,l|l-1}, \tag{6.12}$$

and non-decreasing if

$$v_{s,l-1,\cdot} \geq v_{sl\cdot} \geq \cdots \tag{6.13}$$

and

$$\psi_{s,l-1|l-2} \leq \psi_{s,l|l-1}. \tag{6.14}$$

Proof. From (6.2) we have $(j = b_s + 1, b_s + 2, \ldots)$

$$\begin{aligned}
\psi_{s,j|j-1} - \psi_{s,j+1|j} &= \frac{\varrho^2 \psi_{s,j-1|j-2}\varphi}{v_{s,j-1,\cdot}\psi_{s,j-1|j-2} + \varphi} + (1-\varrho^2)\lambda - \frac{\varrho^2 \psi_{s,j|j-1}\varphi}{v_{sj\cdot}\psi_{s,j|j-1} + \varphi} - (1-\varrho^2)\lambda \\
&= \varrho^2 \varphi \left[\frac{\psi_{s,j-1|j-2}}{v_{s,j-1,\cdot}\psi_{s,j-1|j-2} + \varphi} - \frac{\psi_{s,j|j-1}}{v_{sj\cdot}\psi_{s,j|j-1} + \varphi} \right] \\
&= \frac{\varrho^2 \varphi \left[\psi_{s,j-1|j-2}(v_{sj\cdot}\psi_{s,j|j-1} + \varphi) - \psi_{s,j|j-1}(v_{s,j-1,\cdot}\psi_{s,j-1|j-2} + \varphi) \right]}{(v_{s,j-1,\cdot}\psi_{s,j-1|j-2} + \varphi)(v_{sj\cdot}\psi_{s,j|j-1} + \varphi)}
\end{aligned}$$

giving

$$\psi_{s,j|j-1} - \psi_{s,j+1|j} = \frac{\varrho^2 \varphi \left[\psi_{s,j-1|j-2}\psi_{s,j|j-1}(v_{sj\cdot} - v_{s,j-1,\cdot}) + \varphi(\psi_{s,j-1|j-2} - \psi_{s,j|j-1}) \right]}{(v_{s,j-1,\cdot}\psi_{s,j-1|j-2} + \varphi)(v_{sj\cdot}\psi_{s,j|j-1} + \varphi)}. \quad (6.15)$$

Inserting (6.11) and (6.12) into (6.15) with $j = l$ yields

$$\psi_{s,l|l-1} - \psi_{s,l+1|l} \geq 0$$

which is equivalent to

$$\psi_{s,l|l-1} \geq \psi_{s,l+1|l}.$$

Repeated use of (6.15) and (6.11) with $(j = l+1, l+2, \ldots)$ shows that $\{\psi_{s,j|j-1}\}_{j=l}^{\infty}$ is non-increasing and by including (6.12) we have that $\{\psi_{s,j|j-1}\}_{j=l-1}^{\infty}$ is non-increasing.

Analogously, by using (6.13) instead of (6.11) and (6.14) instead of (6.12) we have from (6.15) that $\{\psi_{s,j|j-1}\}_{j=l-1}^{\infty}$ is non-decreasing if (6.13) and (6.14) hold. □

(6.13) will typically hold for car models which are on their way out of the market. It seems reasonable that the estimation error increases for these car models since the estimation is based on less and less information in each year. Whether (6.14) holds or not for these car models must be checked in each particular situation and it depends on the parameter values and how the risk volume fluctuates from year $l-2$ to year $l-1$. The strict version of (6.14) will not hold if $\varrho = 1$ because in that case (6.2) becomes (with $t = l-1$)

$$\psi_{s,l|l-1} = \psi_{s,l-1|l-2} \frac{\varphi}{v_{s,l-1,\cdot}\psi_{s,l-1|l-2} + \varphi} \leq \psi_{s,l-1|l-2}$$

which contradicts $\psi_{s,l-1|l-2} < \psi_{s,l|l-1}$. That the strict version of (6.14) may be fulfilled in some situations if $\varrho < 1$ is realized by putting $t = l-2$ in (6.2) and letting $v_{s,l-2,\cdot} \to \infty$. Then $\psi_{s,l-1|l-2} \to (1-\varrho^2)\lambda$. Assume $\lambda > 0$ and $0 < \varrho < 1$. By putting $t = l-1$ in (6.2) we then get

$$\psi_{s,l|l-1} = \varrho^2 \frac{\psi_{s,l-1|l-2}\varphi}{v_{s,l-1,\cdot}\psi_{s,l-1|l-2} + \varphi} + (1-\varrho^2)\lambda > (1-\varrho^2)\lambda = \lim_{v_{s,l-2,\cdot} \to \infty} \psi_{s,l-1|l-2}.$$

Hence, for a sufficiently large risk volume in year $l-2$ we have

$$\psi_{s,l|l-1} > \psi_{s,l-1|l-2},$$

which is exactly the strict version of (6.14). In other words if $\lambda > 0$, $0 < \varrho < 1$ and (6.13) is fulfilled then the sequence $\{\psi_{s,j|j-1}\}_{j=l-1}^{\infty}$ is non-decreasing if the risk volume of car model s in year $l-2$ becomes sufficiently large.

An interesting special case of Theorem 6.3 is the following corollary.

Corollary 6.2 *The sequence* $\{\psi_{s,j|j-1}\}_{j=b_s}^{\infty}$ *is non-increasing if*

$$v_{sb_s\cdot} \leq v_{s,b_s+1,\cdot} \leq \cdots. \tag{6.16}$$

Proof. From (6.1) and (6.2) with $t = b_s$ we have

$$\psi_{s,b_s|b_s-1} - \psi_{s,b_s+1|b_s} = \lambda - \varrho^2\lambda\frac{\varphi}{v_{sb_s\cdot}\lambda + \varphi} - \left(1 - \varrho^2\right)\lambda = \varrho^2\lambda\frac{v_{sb_s\cdot}\lambda}{v_{sb_s\cdot}\lambda + \varphi} \geq 0$$

which is the same as

$$\psi_{s,b_s|b_s-1} \geq \psi_{s,b_s+1|b_s}. \tag{6.17}$$

Now using Theorem 6.3 with $l = b_s+1$, (6.16) instead of (6.11), and (6.17) instead of (6.12) prove the corollary. □

Car models recently introduced into the market typically have increasing risk volumes. Corollary 6.2 shows that these car models will have an improvement in estimation error as long as the risk volumes increase. This is reasonable since we then base our estimation on more and more information for each year.

Chapter 7

Time-heterogeneous Model for Two Portfolios

7.1 Introduction

From Theorem 5.2 and Assumption 5.2 we see that in order to compute $\tilde{m}_{s,u+1}$ we have to know the technical variables $\mathbf{t}_{sj}$ $(j = b_s, \ldots, u+1)$. In Storebrand we used, in vehicle damage insurance, the technical variables: engine power, price/weight and a constant term. A problem arises for car models that are no longer sold as new in Norway, since the price will be unknown.

Sundt has proposed the following solutions to this problem:

1. Make an adjustment for inflation of the last known price for a new car of car model s. (See SUNDT(1987a, p. 64) and SUNDT(1987b, p. 194)).

2. Make a subjective assessment of the price for a new car of car model s, based on prices for new cars of similar car models still sold in Norway. (See SUNDT(1987b, p. 194)).

3. Replace the prior estimator in each year by an old credibility estimator. (See SUNDT (1987b)).

All of the three proposals have disadvantages. The price of a car model is not influenced only by inflation. This is not taken into consideration in proposal 1.

A disadvantage of the second proposal is that the method is very time-consuming and expensive, since it necessitates a manual investigation.

The third proposal will not solve this problem satisfactorily in e.g. the model of Chapter 5. Using his notation, SUNDT(1987b) wants to estimate a random variable m and assumes that $\mathrm{E}(m|\Theta) = \mathbf{a}'\mathbf{b}(\Theta)$ where $\mathbf{a}$ is a non-random $q \times 1$ vector and $\mathbf{b}(\Theta)$ is a $q \times 1$ vector function of an unknown random variable Θ. He then considers estimators of the form

$$m^* = \mathbf{a}'[\mathbf{Z}\mathbf{b}^* + (\mathbf{I}_q - \mathbf{Z})\boldsymbol{\beta}], \tag{7.1}$$

where $\boldsymbol{\beta} = \mathrm{E}[\mathbf{b}(\Theta)]$, $\mathbf{Z}$ is a non-random $q \times q$ matrix, and $\mathbf{b}^*$ is a Θ-unbiased estimator of $\mathbf{b}(\Theta)$ (as defined in Definition 10.1). In addition he assumes that he has got an old estimator $\dot{\mathbf{b}}$ of $\mathbf{b}(\Theta)$, and that $\dot{\mathbf{b}}$ is conditionally independent of m and $\mathbf{b}^*$ given Θ. Finally he defines

$$\dot{m} = \mathbf{a}'[\mathbf{Z}\mathbf{b}^* + (\mathbf{I}_q - \mathbf{Z})\dot{\mathbf{b}}],$$

and gives sufficient and necessary conditions for $\dot{m}$ to be a better estimator of m than m^* is.

Unfortunately we can not use this result here. From (5.6) we see that $\tilde{m}_{s,t+1|t}$ is not of the form (7.1).

Even if we should insist on replacing the prior estimators in (5.6) by old estimators we get into trouble. To illustrate this we assume that h_s is the last year we know the price of car model s. By substituting $t = h_s$ into (5.6) we have

$$\tilde{m}_{s,h_s+1|h_s} = \varrho_{h_s}\left[\zeta_{sh_s} Y_{sh_s\cdot} + (1-\zeta_{sh_s})\tilde{m}_{s,h_s|h_s-1} - \mu_{sh_s}\right] + \mu_{s,h_s+1} \tag{7.2}$$

with

$$\zeta_{sh_s} = \frac{\psi_{s,h_s|h_s-1} v_{sh_s\cdot}}{\psi_{s,h_s|h_s-1} v_{sh_s\cdot} + \varphi_{h_s}}.$$

Assume now that the structural parameters are known (or at least are estimated). Since the price of car model s is unknown in year $h_s + 1$, μ_{s,h_s+1} is unknown. This is the only unknown quantity on the right hand side of (7.2). Following the idea in SUNDT(1987b) we want to replace μ_{s,h_s+1} by an old estimator of $m_{s,h_s+1}(\Theta_{s,h_s+1})$. The most recent estimator we are able to calculate is the credibility estimator of $m_{s,h_s+1}(\Theta_{s,h_s+1})$ based on observations up to year $h_s - 1$ inclusive. Substituting $t = h_s$ and $v_{sh_s\cdot} = 0$ into (5.6) we have (putting $v_{sh_s\cdot} = 0$ is equivalent to having no observations from year h_s of car model s, so in this particular case $\tilde{m}_{s,h_s+1|h_s}$ is equal to $\tilde{m}_{s,h_s+1|h_s-1}$)

$$\tilde{m}_{s,h_s+1|h_s-1} = \varrho_{h_s}\left[\tilde{m}_{s,h_s|h_s-1} - \mu_{sh_s}\right] + \mu_{s,h_s+1}.$$

We see that this estimator depends on μ_{s,h_s+1} as well. Thus, replacing μ_{s,h_s+1} in (7.2) by $\tilde{m}_{s,h_s+1|h_s-1}$ does not solve our problem.

The difficulty seems to arise from the fact that in Chapter 5 we are in a dynamic situation whereas in SUNDT(1987b) we are in a static situation (μ_{st} independent of time).

The present chapter will consider a new proposal to the solution of this problem. The idea here is that since the price for a new car of car model s is unknown, we will, for these car models only, incorporate the price into the unobservable risk characteristics Θ_s. We may then replace the price with other observable technical variables.

7.2 The model

We divide the portfolio of car models into two sub-portfolios A and B. They consist of car models still sold as new in Norway and car models no longer for sale in Norway, respectively. Car models may move from sub-portfolio A to sub-portfolio B, but not vice versa (see Figure 7.1). For the sake of simplicity we assume that a car model stays in sub-portfolio A for the whole year in which it is no longer for sale. The transition from sub-portfolio A to sub-portfolio B will then take place at the beginning of the year after. We also assume that the time car model k has spent in sub-portfolio A does not influence the risk characteristics Θ_{kj} prior to transition to sub-portfolio B (in particular this means that we assume the second order moments for car models in sub-portfolio A to be independent of the time the car model has been in sub-portfolio A). This is a bit unrealistic. It seems reasonable to assume that the risk characteristics from car models which has been taken out of the market shortly after it was introduced, follows another distribution than the risk characteristics of the car models which have stayed in sub-portfolio A for a longer period

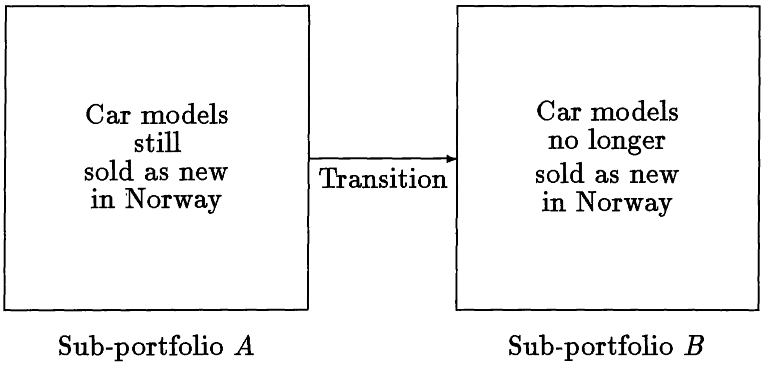

Figure 7.1: How the original portfolio is divided into two sub-portfolios.

of time. It could very well be that the car model was on the market for only a short time, because it was not a good one. Perhaps it had some constructional errors? We introduce this assumption to simplify the estimation of parameters.

The definitions in Section 4.2 are modified in the following way (see Figure 7.2 for a schematic illustration):

$\mathcal{A}_j^p$ = set of indices from all car models still in sub-portfolio p in year j $(p = A, B)$
$\mathcal{A}_{ij}^p$ = $\mathcal{A}_i^p \cap \mathcal{A}_j^p \quad (p = A, B)$
$\mathcal{A}_{ij}^{BA}$ = $\mathcal{A}_i^B \cap \mathcal{A}_j^A \quad (i > j)$
r_k = last year we have observations from car model k in sub-portfolio A
n = number of years the two sub-portfolios have been under observation

$\mathcal{B}_k^{(l)}$, ${}^{i}_{j}K^p$, K_j^p, $n_k^{(l),p}$, $K^{(l),p}$, ${}^{i}_{j}\mathbf{Y}^p_{\Sigma i}$, $\mathbf{Y}^p_{\Sigma i}$, ${}^{i}_{j}\mathbf{T}^p_j$, $\mathbf{T}^p_j$ and $\mathbf{D}^p_j$ have the same definitions as in Chapter 4, but are valid for sub-portfolio p only $(p = A, B)$. (We put $n_s^A = 0$ if car model s starts in sub-portfolio B, that is, if $s \in \mathcal{A}_{b_s}^B$). The other definitions in Section 4.2 are kept unchanged.

The assumptions of Section 4.3 are changed in a similar way to be valid for sub-portfolio p $(p = A, B)$.

The assumptions of Section 5.2 for all $k \in \mathcal{A}_j$ are replaced by (for all $k \in \mathcal{A}_j^p$; $p = A, B$)

7.1. $\mathrm{E}(Y_{kji}|\Theta_{kj}) = m_{kj}(\Theta_{kj})$

7.2. $\mu_{kj} = \mathrm{E}[m_{kj}(\Theta_{kj})] = \mathbf{t}'_{kj}\boldsymbol{\beta}_j^p$

7.3. $\mathrm{Var}(Y_{kji}|\Theta_{kj}) = [s_j^p(\Theta_{kj})]^2/v_{kji}$

7.4. $\mathrm{E}[s_j^p(\Theta_{kj})]^2 = \varphi_j^p$

7.5. $\mathrm{Cov}(m_{k'i}(\Theta_{k'i}), m_{k'j}(\Theta_{k'j})) = \lambda_{ij}^p \quad (k' \in \mathcal{A}_{ij}^p \; ; \; p = A, B) \;, \; \lambda_j^p = \lambda_{jj}^p \quad (p = A, B)$
$\mathrm{Cov}(m_{k'i}(\Theta_{k'i}), m_{k'j}(\Theta_{k'j})) = \lambda_{ij}^{(r_{k'}),BA} \quad (k' \in \mathcal{A}_{ij}^{BA} \; ; \; i > j)$

7.6. $\lambda_{i+1,j}^p = \varrho_i^p \lambda_{ij}^p \quad (p = A, B \; ; \; i \geq j) \qquad \lambda_{i+1,j}^{(i),BA} = \varrho_i^{BA} \lambda_{ij}^A \quad (i \geq j)$
$\lambda_{i+1,j}^{(r_{k'}),BA} = \varrho_i^B \lambda_{ij}^{(r_{k'}),BA} \quad (k' \in \mathcal{A}_{i+1,i}^B \cap \mathcal{A}_j^A \; ; \; i > j)$

7.7. $\mathrm{Corr}(m_{k'i}(\Theta_{k'i}), m_{k'j}(\Theta_{k'j})) = \pi_{ij}^p \in [0,1] \quad (k' \in \mathcal{A}_{ij}^p \; ; \; p = A, B)$
$\mathrm{Corr}(m_{k'i}(\Theta_{k'i}), m_{k'j}(\Theta_{k'j})) = \pi_{ij}^{(r_{k'}),BA} \in [0,1] \quad (k' \in \mathcal{A}_{ij}^{BA} \; ; \; i > j)$

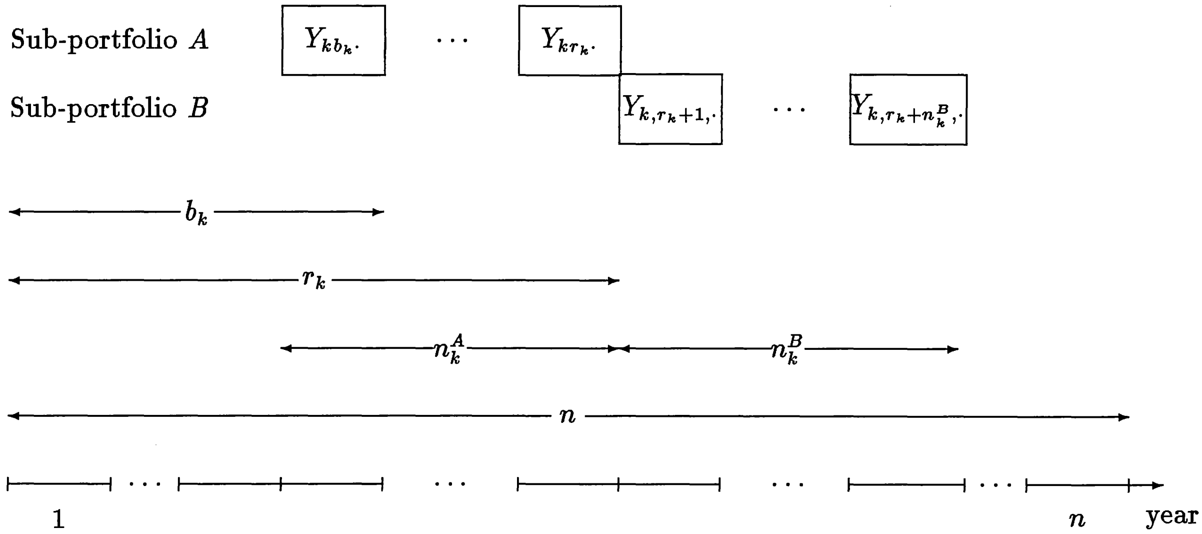

Figure 7.2: Observations and variables associated with car model k going from sub-portfolio A to sub-portfolio B during the observational period (to make the illustration more clear we assume that we have observations from car model k from every year $b_k, \dots, r_k + n_k^B$ even though this is in general not necessary).

Since the Θ_{kj}'s for fixed k contain different information in the two sub-portfolios it is natural that the moments do depend on the sub-portfolio.

By thoroughly comparing the present model with the model in Chapter 5, we see that there are few differences between the two models. For car models starting in sub-portfolio B, the two models will be identical. Car models starting in sub-portfolio A will sooner or later be transferred to sub-portfolio B. Until this happens, the situation is identical to the one in Chapter 5. For a car model going from sub-portfolio A to sub-portfolio B, the definition and dimension of $\mathbf{t}_{kj}$, the vector of technical variables, is altered. Using an idea in HESSELAGER(1989) we can write $\theta_{kj} = \theta(\mathbf{c}_{kj})$ for given $\Theta_{kj} = \theta_{kj}$, $\mathbf{c}_{kj}$ being the vector of the unobservable risk characteristics. When passing from sub-portfolio A to B the price, in our example, is transferred from $\mathbf{t}_{kj}$ to $\mathbf{c}_{kj}$. Therefore we should select a new set of technical variables in $\mathbf{t}_{kj}$ for $k \in \mathcal{A}_j^B$. These variables must of course be observable. They should be selected in such a way that we lose as little prior information as possible. For criteria on how to select these variables, and other problems in finding reasonable technical variables, we refer to SUNDT(1987a, subsection 3.5). The consequence of this is that $\mathbf{t}_{kj}$ and $\mathbf{c}_{kj}$ have new definitions and dimensions for $k \in \mathcal{A}_j^B$ and that the Θ_{kj}'s for fixed k contain different information in the two sub-portfolios, but this does not cause any problems in the model of Chapter 5. In fact, for a specific car model s, the model of the present chapter can be formulated within the model of Chapter 5.

7.3 Recursive credibility estimators

In this model we get the recursive procedure:

Theorem 7.1 *For* $(j = b_s - 1, b_s, \dots)$ *let* $\tilde{m}_{s,j+1|j}$ *be the credibility estimator of* $m_{s,j+1}(\Theta_{s,j+1})$ *based on* $(({}_{r_s}\mathbf{Y}^A_{s\Sigma})', ({}_j\mathbf{Y}^B_{s\Sigma})')'$ $(j \geq r_s)$ $({}_{r_s}\mathbf{Y}^A_{s\Sigma} = \emptyset$ *if* $n_s^A = 0$ *and* ${}_j\mathbf{Y}^B_{s\Sigma} = \emptyset$ *if* $s \notin \mathcal{A}_j^B)$ *and* *define*

$$\psi_{s,j+1|j} = \mathrm{E}[(m_{s,j+1}(\Theta_{s,j+1}) - \tilde{m}_{s,j+1|j})^2].$$

To improve readability we introduce ($s \in \mathcal{A}_j^p$; $p = A, B$; $j = b_s, b_s + 1, \ldots$)

$$\begin{aligned} V_{sj} &= \frac{v_{sj}.\psi_{s,j|j-1}}{v_{sj}.\psi_{s,j|j-1} + \varphi_j^p} Y_{sj}. + \frac{\varphi_j^p}{v_{sj}.\psi_{s,j|j-1} + \varphi_j^p} \tilde{m}_{s,j|j-1} - \mu_{sj} \\ W_{sj} &= \frac{\psi_{s,j|j-1}\varphi_j^p}{v_{sj}.\psi_{s,j|j-1} + \varphi_j^p} - \lambda_j^p. \end{aligned}$$

Then we have

$$\tilde{m}_{sb_s} = \mu_{sb_s} \qquad \psi_{sb_s} = \lambda_{b_s}^p.$$

Further, for $s \in \mathcal{A}_{b_s}^A$ *we have*

$$\begin{aligned} \psi_{s,j+1|j} &= \left(\varrho_j^A\right)^2 W_{sj} + \lambda_{j+1}^A \quad (j = b_s, \ldots, r_s - 1) \\ \psi_{s,r_s+1|r_s} &= \left(\varrho_{r_s}^{BA}\right)^2 W_{sr_s} + \lambda_{r_s+1}^B \\ \tilde{m}_{s,j+1|j} &= \varrho_j^A V_{sj} + \mu_{s,j+1} \quad (j = b_s, \ldots, r_s - 1) \\ \tilde{m}_{s,r_s+1|r_s} &= \varrho_{r_s}^{BA} V_{sr_s} + \mu_{s,r_s+1} \end{aligned}$$

and for $s \in \mathcal{A}_{b_s}^p$ ($p = A, B$) *we have*

$$\begin{aligned} \psi_{s,j+1|j} &= \left(\varrho_j^B\right)^2 W_{sj} + \lambda_{j+1}^B \\ \tilde{m}_{s,j+1|j} &= \varrho_j^B V_{sj} + \mu_{s,j+1} \end{aligned}$$

where ($j = r_s + 1, r_s + 2, \ldots$) *if* $s \in \mathcal{A}_{b_s}^A$ *and* ($j = b_s, b_s + 1, \ldots$) *if* $s \in \mathcal{A}_{b_s}^B$.

Proof. Since s is fixed the theorem follows from Theorem 5.2. □

Chapter 8

Connection with the Kalman Filter

8.1 Introduction

In the present chapter we are going to investigate the connection between the Kalman filter and the time-heterogeneous model of Chapter 5. An analogous connection with the Kalman filter can be found to exist for the models of Chapters 6 and 7 as well.

The recursive credibility estimators and their estimation errors (Theorem 5.2) are strongly connected with the "one-step ahead" predictions and their prediction errors in the (linear) Kalman filter. This is not surprising. In our set-up we want to estimate an unobservable random vector by an optimal linear estimator using a quadratic loss function. In the Kalman filter set-up we want to estimate an unobservable random vector in a linear model by an optimal estimator using a quadratic loss function. To make this relation more clear we will prove Theorem 5.2 by the use of the Kalman filter.

The connection between credibility theory and the Kalman filter has earlier been investigated by e.g. deJong & Zehnwirth(1983).

This chapter will closely follow Neuhaus(1989). But since it is sufficient for our use, we will restrict ourselves to the case of one-dimensional observations and state vectors.

8.2 Kalman filter

Let $\{y_t\}_{t=1}^{\infty}$ be a sequence of random elements. We assume that y_t is linearly regressed on an unobserved random state element x_t. Thus the observations are connected to the state vector through the observation equation

$$y_t = F_t x_t + v_t.$$

Assume further that the state elements evolve in accordance with the system equation

$$x_t = G_t x_{t-1} + u_t + w_t.$$

The sequence $\{u_t\}_{t=1}^{\infty}$ is assumed to consist of non-random known model parameters; F_t and G_t are assumed to be non-random and known, v_t and w_t are assumed to be random, and all v_t and w_t are mutually independent with mean zero and known variances $\mathrm{Var}(v_t) = Q_t$ and $\mathrm{Var}(w_t) = R_t$. Assume that our objective is to estimate the state elements $\{x_t\}_{t=1}^{\infty}$ based on the observations $\{y_t\}_{t=1}^{\infty}$, and that the prior mean $\mu_0 = \mathrm{E}(x_0)$ and variance $\psi_0 = \mathrm{Var}(x_0)$

are both known. Let $\bar{x}_{s|t}$ ($\bar{y}_{s|t}$) be the optimal linear estimator of x_s (y_s) using a quadratic loss function, given $(y_1, \ldots, y_t)'$, and let

$$L^x_{s|t} = \mathrm{E}(x_s - \bar{x}_{s|t})^2 \quad (L^y_{s|t} = \mathrm{E}(y_s - \bar{y}_{s|t})^2)$$

be the estimation errors. The linear Kalman filter is defined by the following recursive formulae.

Theorem 8.1 *The Kalman filter is initialized by*

$$\bar{x}_{0|0} = \mu_0 \qquad L^x_{0|0} = \psi_0.$$

For $t = 1, 2, \ldots$ *we have*

$$\begin{aligned}
\bar{x}_{t|t-1} &= G_t\bar{x}_{t-1|t-1} + u_t \\
L^x_{t|t-1} &= R_t + G_t^2 L^x_{t-1|t-1} \\
\bar{y}_{t|t-1} &= F_t\bar{x}_{t|t-1} \\
L^y_{t|t-1} &= Q_t + F_t^2 L^x_{t|t-1} \\
\bar{x}_{t|t} &= \bar{x}_{t|t-1} + K_t\left(y_t - \bar{y}_{t|t-1}\right) \\
L^x_{t|t} &= (1 - K_tF_t)L^x_{t|t-1}
\end{aligned}$$

where we have defined the "Kalman gain"

$$K_t = L^x_{t|t-1}F_t\left(L^y_{t|t-1}\right)^{-1}.$$

Proof. See e.g. NEUHAUS(1989, Theorem 2.11). □

8.3 Time-heterogeneous model in a Kalman filter setting

We are now able to put the time-heterogeneous model into a Kalman filter framework. Consider the observation equation

$$Y_{kji} = m_{kj}(\Theta_{kj}) + \varepsilon_{kji}, \tag{8.1}$$

and the system equation

$$m_{k,j+1}(\Theta_{k,j+1}) = \varrho_j m_{kj}(\Theta_{kj}) + \mu_{k,j+1} - \varrho_j\mu_{kj} + \omega_{k,j+1}. \tag{8.2}$$

Here $\varepsilon_{kj1}, \ldots, \varepsilon_{kjI_{kj}}$ are stochastically independent of Θ_{kj}, $\omega_{k,j+1}$ is independent of $\Theta_{kb_k}, \ldots, \Theta_{kj}$ $(j = b_k, \ldots)$, and all the ε_{kji}'s and the ω_{kj}'s are mutually independent. We let $\mathrm{E}(\varepsilon_{kji}) = 0$, $\mathrm{Var}(\varepsilon_{kji}) = \varphi_j/v_{kji}$, $\mathrm{E}(\omega_{k,j+1}) = 0$, $\mathrm{Var}(\omega_{k,j+1}) = \lambda_{j+1} - \varrho_j^2\lambda_j$ and $\mathrm{Cov}(m_{ki}(\Theta_{ki}), m_{kj}(\Theta_{kj})) = \lambda_{ij}$. Then the model described by (8.1) and (8.2) has the same first and second order moments structure as the model in Chapter 5. From (2.6) we see that the credibility estimator only depends on the first and second order moments of the model. Thus the optimal linear estimator in the two models must be the same. We have also shown (Theorem 5.1) that the optimal linear estimator based on all of the original observations ${}_t\mathbf{Y}$ is identical to the optimal linear estimator based on the summarized observations ${}_t\mathbf{Y}_{s\Sigma}$ from car model s.

Therefore we can restrict ourselves to studying the summarized observations, and replace (8.1) by the observation equation

$$Y_{kj\cdot} = m_{kj}(\Theta_{kj}) + \varepsilon_{kj},$$

where ε_{kj} are stochastically independent of Θ_{kj}, $\mathrm{E}(\varepsilon_{kj}) = 0$, and $\mathrm{Var}(\varepsilon_{kj}) = \varphi_j/v_{kj\cdot}$. In addition, we now assume that all the ε_{kj}'s and the ω_{kj}'s are mutually independent.

Before continuing we are going to show an identity which we will use below. Consider

$$\begin{aligned}
\mathrm{E}(Y_{kt\cdot} - \tilde{m}_{k,t|t-1})^2 &= \mathrm{E}(Y_{kt\cdot} - m_{kt}(\Theta_{kt}) + m_{kt}(\Theta_{kt}) - \tilde{m}_{k,t|t-1})^2 \\
&= \mathrm{E}(Y_{kt\cdot} - m_{kt}(\Theta_{kt}))^2 + \mathrm{E}(m_{kt}(\Theta_{kt}) - \tilde{m}_{k,t|t-1})^2 \\
&\quad + 2\mathrm{E}(Y_{kt\cdot} - m_{kt}(\Theta_{kt}))(m_{kt}(\Theta_{kt}) - \tilde{m}_{k,t|t-1}) \\
&\overset{(1)}{=} \mathrm{E}\{\mathrm{E}[(Y_{kt\cdot} - m_{kt}(\Theta_{kt}))^2|\Theta_{kt}]\} + \mathrm{E}(m_{kt}(\Theta_{kt}) - \tilde{m}_{k,t|t-1})^2 \\
&\quad + 2\mathrm{E}\{\mathrm{E}[(Y_{kt\cdot} - m_{kt}(\Theta_{kt}))(m_{kt}(\Theta_{kt}) - \tilde{m}_{k,t|t-1})|\Theta_k]\} \\
&\overset{(2)}{=} \mathrm{E}[\mathrm{Var}(Y_{kt\cdot}|\Theta_{kt})] + \psi_{k,t|t-1} + 0
\end{aligned}$$

where we in (1) have used (A.15) and in (2) we have used the definition of conditional variance, the definition of $\psi_{k,t|t-1}$, and that $\tilde{m}_{k,t|t-1}$ is a linear combination of $Y_{kb_k\cdot}, \ldots, Y_{k,t-1,\cdot}$ which are independent of $Y_{kt\cdot}$ for given Θ_k (cf. Assumption 4.2). By substituting (A.6) and Assumption 5.4 we obtain

$$\mathrm{E}(Y_{kt\cdot} - \tilde{m}_{k,t|t-1})^2 = \frac{\varphi_t}{v_{kt\cdot}} + \psi_{k,t|t-1}. \tag{8.3}$$

Using the notation of Theorem 8.1 we see that we have

$$\begin{array}{lllll}
y_t = Y_{kt\cdot} & F_t = 1 & x_t = m_{kt}(\Theta_{kt}) & v_t = \varepsilon_{kt} & Q_t = \varphi_t/v_{kt\cdot} \\
G_t = \varrho_{t-1} & u_t = \mu_{kt} - \varrho_{t-1}\mu_{k,t-1} & w_t = \omega_{kt} & R_t = \lambda_t - \varrho_{t-1}^2\lambda_{t-1}.
\end{array}$$

Further,

$$\begin{aligned}
\bar{x}_{t|t-1} &= \bar{y}_{t|t-1} = \tilde{m}_{k,t|t-1} \\
L^x_{t|t-1} &= \mathrm{E}(x_t - \bar{x}_{t|t-1})^2 = \mathrm{E}(m_{kt}(\Theta_{kt}) - \tilde{m}_{k,t|t-1})^2 = \psi_{k,t|t-1} \\
L^y_{t|t-1} &= \mathrm{E}(y_t - \bar{y}_{t|t-1})^2 = \mathrm{E}(Y_{kt\cdot} - \tilde{m}_{k,t|t-1})^2 \overset{(1)}{=} \psi_{k,t|t-1} + \varphi_t/v_{kt\cdot} \\
K_t &= L^x_{t|t-1}F_t[L^y_{t|t-1}]^{-1} = \psi_{k,t|t-1}/(\psi_{k,t|t-1} + \varphi_t/v_{kt\cdot}) = v_{kt\cdot}\psi_{k,t|t-1}/(v_{kt\cdot}\psi_{k,t|t-1} + \varphi_t)
\end{aligned}$$

with (1) being a consequence of (8.3). From Theorem 8.1 we have that the Kalman filter ("one-step ahead" prediction) is given by ($b_k - 1$ corresponds to year 0 in Theorem 8.1):

Initialization:

$$\begin{aligned}
\tilde{m}_{k,b_k|b_k-1} &= \bar{x}_{b_k|0} = \varrho_{b_k-1}\bar{x}_{0|0} + \mu_{kb_k} - \varrho_{b_k-1}\mu_{k,b_k-1} = \mu_{kb_k} \\
\psi_{k,b_k|b_k-1} &= L^x_{b_k|0} = \lambda_{b_k} - \varrho^2_{b_k-1}\lambda_{b_k-1} + \varrho^2_{b_k-1}L^x_{0|0} = \lambda_{b_k}
\end{aligned}$$

For ($t = b_k, b_k+1, \ldots$):

$$\begin{aligned}
\psi_{k,t+1|t} &= L^x_{t+1|t} = R_{t+1} + G^2_{t+1}(1 - K_tF_t)L^x_{t|t-1} \\
&= \lambda_{t+1} - \varrho_t^2\lambda_t + \varrho_t^2\left(1 - \frac{v_{kt\cdot}\psi_{k,t|t-1}}{v_{kt\cdot}\psi_{k,t|t-1} + \varphi_t}1\right)\psi_{k,t|t-1}
\end{aligned}$$

$$
\begin{aligned}
&= \varrho_t^2 \left(\frac{\psi_{k,t|t-1}\varphi_t}{v_{kt\cdot}\psi_{k,t|t-1} + \varphi_t} - \lambda_t \right) + \lambda_{t+1} \\
\tilde{m}_{k,t+1|t} &= \bar{x}_{t+1|t} = G_{t+1}[\bar{x}_{t|t-1} + K_t(y_t - \bar{y}_{t|t-1})] + u_{t+1} \\
&= \varrho_t \left[\tilde{m}_{k,t|t-1} + \frac{v_{kt\cdot}\psi_{k,t|t-1}}{v_{kt\cdot}\psi_{k,t|t-1} + \varphi_t}(Y_{kt\cdot} - \tilde{m}_{k,t|t-1}) \right] + \mu_{k,t+1} - \varrho_t\mu_{kt} \\
&= \varrho_t \left[\frac{v_{kt\cdot}\psi_{k,t|t-1}}{v_{kt\cdot}\psi_{k,t|t-1} + \varphi_t} Y_{kt\cdot} + \frac{\varphi_t}{v_{kt\cdot}\psi_{k,t|t-1} + \varphi_t}\tilde{m}_{k,t|t-1} - \mu_{kt} \right] + \mu_{k,t+1},
\end{aligned}
$$

which is Theorem 5.2.

We see that recursive credibility estimation is the same *method* as Kalman-filtering, but they are derived under different model assumptions.

Corollary 5.1 is an easy consequence of the Kalman filter. This is realized by observing that $\bar{x}_{t|t} = \tilde{m}_{k,t|t}$, $K_t = \zeta_{kt}$, $L^x_{t|t} = \psi_{k,t|t}$ in addition to the quantities given above. Inserting these quantities into Theorem 8.1 yields Corollary 5.1.

Chapter 9

Connection with Other Related Models

9.1 The Bühlmann-Straub model

By putting $\mu_k \equiv \mu$ and $\Theta_{kj} = \Theta_{kb_k}$ $\forall j$ (implying $\varrho = 1$) in the model of Chapter 6 we get the model in BÜHLMANN & STRAUB(1970). So Corollary 6.1 gives a recursive procedure for computing the credibility estimators and their estimation errors in this model as well.

9.2 The generalized Bühlmann-Straub model of Sundt

For each car model in the model of Chapter 6 we are in fact using the model in subsection 4.2.3.2 in SUNDT(1981). In the model of Chapter 5 we are for each car model using a special case of the models in subsections 4.1 and 4.2.1.1 in the same paper. We can do this without losing any information that lies in the original observations, due to our model assumptions and the use of linear sufficiency. The time-heterogeneous model in SUNDT(1983) is again a special case of the model in subsection 4.2.1.1 in SUNDT(1981).

9.3 The non-hierarchical model in Sundt(1987a)

There are at least two ways of making a relationship between the model of Chapter 6 and the non-hierarchical model in SUNDT(1987a).

First, by using observations from one year only in both models (letting $n = 1$), the two models coincide.

In the model in SUNDT(1987a) the risk levels are independent of time. In the model of Chapter 6 the unobservable risk characteristics vary from year to year and therefore the risk levels vary. This is the only difference between the two models (strictly speaking SUNDT(1987a) assumes that the Θ_{kj}'s are identically distributed for fixed j, using the notation of our models, but this assumption can be relaxed in the same way as we have done in our models). So letting $\Theta_{kj} = \Theta_{kb_k}$ be independent of j in the model of Chapter 6 (this implies that $\varrho = 1$), the two models coincide.

The second approach clearly illustrates the difference between the models in the present paper and the model in SUNDT(1987a). He describes a static situation since he wants to estimate the same random variable (the risk level of a car model) from year to year. On the other hand, we describe a dynamic situation in our models since we want to estimate a sequence of random variables (the risk levels of a car model) developing over time.

Chapter 10

Estimation of Parameters

10.1 Introduction

In practical applications the structural parameters will be unknown and have to be estimated.

For the models in Chapters 5 and 7 we see from Theorems 5.2 and 7.1 that we have to estimate parameters both from years from which we have observations and from which we do not have observations. These two cases will be handled separately. We shall call parameters from the first category (year $\leq n$) *past parameters.* These parameters can be estimated using the available observations. Parameters from the last category (year $= n+1$) will be called *future parameters.* From year $n+1$ we have not yet received any claim data. Thus, to be able to estimate the future parameters we must introduce more structure into the models. This will be done by using the theory of martingales. The reason for using this theory is that conditional unbiased estimators of the future parameters given the past parameters are easy to calculate and interpret. Estimation of future parameters is not necessary in the model in Chapter 6, since in that model the parameters are independent of time.

Since $\mu_{kj} = \mathbf{t}'_{kj}\boldsymbol{\beta}^p_j$ $(k \in \mathcal{A}^p_j\ ;\ j = 1,\ldots,n+1)$ and the $\mathbf{t}_{kj}$'s are known, the problem of estimating μ_{kj} $(k \in \mathcal{A}^p_j\ ;\ j = 1,\ldots,n+1)$ is equivalent to that of estimating $\boldsymbol{\beta}^p_j$ $(j = 1,\ldots,n+1)$.

We will not try to find optimal estimators of all the parameters, but merely reasonable estimators which are feasible in practice. The estimators of $\boldsymbol{\beta}^p_j$ may be considered as optimal in some respect, whereas the estimators of φ^p_j, λ^p_j and ϱ^p_j may not. The method of estimating the two latter is the following. We will use the estimation of λ^p_j as an example. ϱ^p_j will be estimated using the same idea.

Assume that we want to base the estimation of λ^p_j on the observations $\mathbf{X}$. We start by finding an intuitively reasonable statistic $f(\mathbf{X})$, say. We calculate $\mathrm{E}[f(\mathbf{X})]$. This expectation will be a linear function of φ^p_j and λ^p_j, that is, $\mathrm{E}[f(\mathbf{X})] = a\varphi^p_j + b\lambda^p_j$ where a and b depend on the risk volumes and the technical variables of the car models. An estimator of λ^p_j is therefore given by $\check{\lambda}^p_j = (f(\mathbf{X}) - a\hat{\varphi}^p_j)/b$, which will be unbiased if $\hat{\varphi}^p_j$ is.

For the estimation of λ^p_j we will also derive an alternative method. This method starts with a statistic which depends on the structural parameters λ^p_j and φ^p_j. We denote this statistic by $f(\mathbf{X}, \lambda^p_j, \varphi^p_j)$ and it will be called a pseudo-statistic (pseudo because the statistic depends on the unknown structural parameters). The expectation of this statistic will be on the form $\mathrm{E}[f(\mathbf{X}, \lambda^p_j, \varphi^p_j)] = a(\lambda^p_j, \varphi^p_j)\varphi^p_j + \lambda^p_j b(\lambda^p_j, \varphi^p_j)$. A pseudo-estimator of λ^p_j is given by

$$\check{\check{\lambda}}_j^p = [f(\mathbf{X}, \lambda_j^p, \varphi_j^p) - a(\lambda_j^p, \varphi_j^p)\hat{\varphi}_j^p]/b(\lambda_j^p, \varphi_j^p), \tag{10.1}$$

which is unbiased if $\hat{\varphi}_j^p$ is. A genuine estimator is obtained by replacing λ_j^p and φ_j^p in (10.1) by one of their estimators, that is,

$$\check{\lambda}_j^p = [f(\mathbf{X}, \hat{\lambda}_j^p, \hat{\varphi}_j^p) - a(\hat{\lambda}_j^p, \hat{\varphi}_j^p)\hat{\varphi}_j^p]/b(\hat{\lambda}_j^p, \hat{\varphi}_j^p).$$

This estimator is usually biased.

Finally, since $\check{\lambda}_j^p$ may attain negative values whereas $\lambda_j^p \geq 0$ we estimate λ_j^p either by

$$\hat{\lambda}_j^p = \max(0, \check{\lambda}_j^p)$$

or

$$\hat{\lambda}_j^p = \begin{cases} \check{\lambda}_j^p & \text{if } \check{\lambda}_j^p > 0 \\ \varepsilon / K_j^p & \text{if } \check{\lambda}_j^p \leq 0 \end{cases}$$

for $\varepsilon > 0$ "small".

The time-heterogeneous model and the time-heterogeneous model for two portfolios will be treated simultaneously, by using the notation from the latter. To obtain the estimators in the former model we put $p = A$ in the latter and delete A from the expressions because of redundant notation.

Throughout this chapter we assume that $p \in \{A, B, BA\}$ is fixed ($p = BA$ for the estimation of ϱ_j^{BA} only) and that all $\mathbf{T}_j^p$ have full rank $q_j^p \leq K_j^p$.

Although $\mathcal{A}_j^p$ denotes the set of indices from all car models still in sub-portfolio p in year j we will in this chapter allow a relaxation of this definition. We may be interested in basing the estimation of the parameters on a sample of car models taken from the sub-portfolios. The reason for why we should want to do this will become clear in Chapter 12.

10.2 Estimation of past parameters

In this section we will propose estimators of the parameters

$$\beta_1^p, \ldots, \beta_n^p, \varphi_1^p, \ldots, \varphi_n^p, \varrho_1^p, \ldots, \varrho_{n-1}^p \text{ and } \lambda_1^p, \ldots, \lambda_n^p.$$

10.2.1 Estimators of φ_j^p

In the present subsection we will make use of the concept Θ-unbiasedness. Our definition is taken from SUNDT(1991b).

Definition 10.1 Let $r(\Theta)$ be a real-valued function of Θ and $\hat{r}$ an estimator of $r(\Theta)$. We say that $\hat{r}$ is a Θ-*unbiased estimator* of $r(\Theta)$ if

$$\mathrm{E}(\hat{r}|\Theta) = r(\Theta) \quad \text{a.s.}$$

Time-heterogeneous models

In order to derive an estimator of φ_j^p ($j = 1, \ldots, n$) we will first find a Θ_{kj}-unbiased estimator of $[s_j^p(\Theta_{kj})]^2$ ($k \in \mathcal{A}_j^p$; $j = 1, \ldots, n$). Assumption 7.3 yields

$$[s_j^p(\Theta_{kj})]^2 = v_{kji}\mathrm{Var}(Y_{kji}|\Theta_{kj}) = v_{kji}\mathrm{E}[(Y_{kji}|\Theta_{kj}) - \mathrm{E}(Y_{kji}|\Theta_{kj})]^2$$

thus, it is natural to base an estimator of $[s_j^p(\Theta_{kj})]^2$ on the statistic

$$(Y_{kji} - Y_{kj\cdot})^2$$

for given Θ_{kj} $(i = 1, \ldots, I_{kj}\ ;\ k \in \mathcal{A}_j^p\ ;\ j = 1, \ldots, n)$. Therefore, we consider

$$\begin{aligned}
\mathrm{E}[(Y_{kji} - Y_{kj\cdot})^2 | \Theta_{kj}] &\overset{(1)}{=} \mathrm{Var}(Y_{kji} - Y_{kj\cdot} | \Theta_{kj}) \\
&= \mathrm{Var}(Y_{kji}|\Theta_{kj}) + \mathrm{Var}(Y_{kj\cdot}|\Theta_{kj}) - 2\mathrm{Cov}(Y_{kji}, Y_{kj\cdot}|\Theta_{kj}) \\
&\overset{(2)}{=} \frac{[s_j^p(\Theta_{kj})]^2}{v_{kji}} + \frac{[s_j^p(\Theta_{kj})]^2}{v_{kj\cdot}} - 2\frac{[s_j^p(\Theta_{kj})]^2}{v_{kj\cdot}}
\end{aligned}$$

where we in (1) have used the fact that $\mathrm{E}(Y_{kji}|\Theta_{kj}) = \mathrm{E}(Y_{kj\cdot}|\Theta_{kj})$, established by comparing (A.16) with Assumption 7.1. (2) is a consequence of Assumption 7.3, (A.6) and (A.5). This yields

$$\mathrm{E}[(Y_{kji} - Y_{kj\cdot})^2|\Theta_{kj}] = [s_j^p(\Theta_{kj})]^2 \left(\frac{v_{kj\cdot} - v_{kji}}{v_{kj\cdot} v_{kji}}\right). \tag{10.2}$$

We therefore have the following Θ_{kj}-unbiased estimator of $[s_j^p(\Theta_{kj})]^2$:

$$\breve{\varphi}_{kj}^p = \frac{1}{I_{kj} - 1} \sum_{i=1}^{I_{kj}} v_{kji}(Y_{kji} - Y_{kj\cdot})^2. \tag{10.3}$$

The Θ_{kj}-unbiasedness is checked by considering (use (10.2) and (10.3))

$$\begin{aligned}
\mathrm{E}(\breve{\varphi}_{kj}^p|\Theta_{kj}) &= \frac{1}{I_{kj} - 1} \sum_{i=1}^{I_{kj}} v_{kji} \mathrm{E}[(Y_{kji} - Y_{kj\cdot})^2|\Theta_{kj}] \\
&= \frac{1}{I_{kj} - 1} \sum_{i=1}^{I_{kj}} v_{kji} [s_j^p(\Theta_{kj})]^2 \left(\frac{v_{kj\cdot} - v_{kji}}{v_{kj\cdot} v_{kji}}\right) \\
&= \frac{1}{(I_{kj} - 1)v_{kj\cdot}} [s_j^p(\Theta_{kj})]^2 \sum_{i=1}^{I_{kj}} (v_{kj\cdot} - v_{kji}) \\
&= \frac{1}{(I_{kj} - 1)v_{kj\cdot}} [s_j^p(\Theta_{kj})]^2 (I_{kj} v_{kj\cdot} - v_{kj\cdot}) \\
&= \frac{1}{(I_{kj} - 1)v_{kj\cdot}} [s_j^p(\Theta_{kj})]^2 (I_{kj} - 1) v_{kj\cdot} = [s_j^p(\Theta_{kj})]^2.
\end{aligned}$$

Since $\varphi_j^p = \mathrm{E}[s_j^p(\Theta_{kj})]^2$, we have that

$$\hat{\varphi}_j^p = \sum_{k \in \mathcal{A}_j^p} u_{kj}^{p,\varphi} \breve{\varphi}_{kj}^p \tag{10.4}$$

where $\sum_{k \in \mathcal{A}_j^p} u_{kj}^{p,\varphi} = 1$, is an unbiased estimator of φ_j^p $(j = 1, \ldots, n)$. $\breve{\varphi}_{kj}^p$ is given by (10.3). A reasonable choice of weights may e.g. be

$$u_{kj}^{p,\varphi} = \frac{(I_{kj} - 1)}{(\sum_{r \in \mathcal{A}_j^p} I_{rj} - K_j^p)} \quad \text{or} \quad u_{kj}^{p,\varphi} = \frac{1}{K_j^p}. \tag{10.5}$$

BÜHLMANN & STRAUB(1970) used the weight $1/K_j^p$ in their analogous estimator. A discussion of both weights in (10.5) is given in SUNDT(1987a, p. 46).

Time-homogeneous model

Since the time-homogeneous model is a special case of the time-heterogeneous model we can use the results from the previous subsection to find an estimator of φ. But since $\varphi_j \equiv \varphi$ we should base our estimator of φ on observations from all previous years. Since $\varphi = \mathrm{E}[s^2(\Theta_{kj})]$ we have that

$$\hat{\varphi} = \sum_{j=1}^{n} \sum_{k \in \mathcal{A}_j} u_{kj}^{\varphi} \check{\varphi}_{kj}, \tag{10.6}$$

where $\sum_{j=1}^{n} \sum_{k \in \mathcal{A}_j} u_{kj}^{\varphi} = 1$ and $\check{\varphi}_{kj}$ is given by (10.3), is an unbiased estimator of φ.

10.2.2 Estimators of β_j^p

Time-heterogeneous models

It is well known (see e.g. NORBERG(1982, p. 76)) that the optimal (with respect to minimizing the variance of the estimator) linear unbiased estimator of β_j^p based on the observations $\mathbf{Y}_{\Sigma j}^p$ is given by

$$\hat{\beta}_j^{p,\mathrm{opt}} = \left\{ (\mathbf{T}_j^p)' \left[\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p) \right]^{-1} \mathbf{T}_j^p \right\}^{-1} (\mathbf{T}_j^p)' \left[\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p) \right]^{-1} \mathbf{Y}_{\Sigma j}^p. \tag{10.7}$$

We have from (A.9)

$$\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p)^{(K_j^p \times K_j^p)} = \mathrm{diag} \left\{ \left(\frac{\varphi_j^p}{v_{kj\cdot}} + \lambda_j^p \right)_{\forall k \in \mathcal{A}_j^p} \right\}$$

implying

$$\left[\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p) \right]^{-1} = \mathrm{diag} \left\{ \left(\frac{v_{kj\cdot}}{\varphi_j^p + v_{kj\cdot} \lambda_j^p} \right)_{\forall k \in \mathcal{A}_j^p} \right\}. \tag{10.8}$$

By putting

$$\mathbf{D}_j^{p,\kappa} = \mathrm{diag} \left\{ \left(\frac{v_{kj\cdot}}{v_{kj\cdot} + \kappa_j^p} \right)_{\forall k \in \mathcal{A}_j^p} \right\}, \tag{10.9}$$

with $\kappa_j^p = \varphi_j^p / \lambda_j^p$, we may rewrite (10.7) as

$$\hat{\beta}_j^{p,\mathrm{opt}} = \left\{ (\mathbf{T}_j^p)' \mathbf{D}_j^{p,\kappa} \mathbf{T}_j^p \right\}^{-1} (\mathbf{T}_j^p)' \mathbf{D}_j^{p,\kappa} \mathbf{Y}_{\Sigma j}^p. \tag{10.10}$$

From (10.9) we see that (10.10) depends on the unknown parameters φ_j^p and λ_j^p. To avoid this problem we replace φ_j^p and λ_j^p by their estimators $\hat{\varphi}_j^p$ and $\hat{\lambda}_j^p$, respectively. This leads to the estimator

$$\hat{\beta}_j^p = \left((\mathbf{T}_j^p)' \mathbf{D}_j^{p,\hat{\kappa}} \mathbf{T}_j^p \right)^{-1} (\mathbf{T}_j^p)' \mathbf{D}_j^{p,\hat{\kappa}} \mathbf{Y}_{\Sigma j}^p \tag{10.11}$$

with

$$\mathbf{D}_j^{p,\hat{\kappa},(K_j^p \times K_j^p)} = \mathrm{diag} \left\{ \left(\frac{v_{kj\cdot}}{v_{kj\cdot} + \hat{\kappa}_j^p} \right)_{\forall k \in \mathcal{A}_j^p} \right\}$$

with $\hat{\kappa}_j^p = \hat{\varphi}_j^p / \hat{\lambda}_j^p$. $\hat{\varphi}_j^p$ is given by (10.4) and $\hat{\lambda}_j^p$ is some estimator of λ_j^p.

An unbiased estimator of β_j^p based on $\mathbf{Y}_{\Sigma j}^p$, which does not depend on estimators of any of the other structural parameters, is

$$\dot{\beta}_j^p = \left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{Y}_{\Sigma j}^p. \quad (j = 1, \ldots, n) \tag{10.12}$$

The unbiasedness is a consequence of

$$\mathrm{E}(\mathbf{Y}_{\Sigma j}^p) = \mathbf{T}_j^p\beta_j^p.$$

We are going to show that among all estimators $\ddot{\beta}_j^p$ of β_j^p,

$$\breve{\beta}_{j,\mathbf{W}}^p = \left((\mathbf{T}_j^p)'\mathbf{W}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{W}\mathbf{Y}_{\Sigma j}^p \tag{10.13}$$

is the one minimizing the function

$$r_{j,\mathbf{W}}^p\left(\ddot{\beta}_j^p\right) = \left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\ddot{\beta}_j^p\right)'\mathbf{W}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\ddot{\beta}_j^p\right)$$

where $\mathbf{W}^{(K_j^p \times K_j^p)}$ is a symmetric positive definite weighting matrix.

Consider first

$$\begin{aligned}
r_{j,\mathbf{W}}^p\left(\ddot{\beta}_j^p\right) &= \left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p + \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p - \mathbf{T}_j^p\ddot{\beta}_j^p\right)'\mathbf{W} \\
&\quad \cdot \left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p + \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p - \mathbf{T}_j^p\ddot{\beta}_j^p\right) \\
&= \left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p\right)'\mathbf{W}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p\right) \\
&\quad + 2\left(\breve{\beta}_{j,\mathbf{W}}^p - \ddot{\beta}_j^p\right)'(\mathbf{T}_j^p)'\mathbf{W}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p\right) \\
&\quad + \left(\breve{\beta}_{j,\mathbf{W}}^p - \ddot{\beta}_j^p\right)'(\mathbf{T}_j^p)'\mathbf{W}\mathbf{T}_j^p\left(\breve{\beta}_{j,\mathbf{W}}^p - \ddot{\beta}_j^p\right).
\end{aligned}$$

Now, we have

$$\begin{aligned}
(\mathbf{T}_j^p)'\mathbf{W}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p\right) &\overset{(1)}{=} (\mathbf{T}_j^p)'\mathbf{W}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{W}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{W}\mathbf{Y}_{\Sigma j}^p\right) \\
&= (\mathbf{T}_j^p)'\mathbf{W}\mathbf{Y}_{\Sigma j}^p - (\mathbf{T}_j^p)'\mathbf{W}\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{W}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{W}\mathbf{Y}_{\Sigma j}^p = \mathbf{0}
\end{aligned}$$

where we in (1) have inserted (10.13). This yields

$$\begin{aligned}
r_{j,\mathbf{W}}^p\left(\ddot{\beta}_j^p\right) &= \left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p\right)'\mathbf{W}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\breve{\beta}_{j,\mathbf{W}}^p\right) \\
&\quad + \left(\breve{\beta}_{j,\mathbf{W}}^p - \ddot{\beta}_j^p\right)'(\mathbf{T}_j^p)'\mathbf{W}\mathbf{T}_j^p\left(\breve{\beta}_{j,\mathbf{W}}^p - \ddot{\beta}_j^p\right),
\end{aligned}$$

and $r_{j,\mathbf{W}}^p$ is minimized for $\ddot{\beta}_j^p = \breve{\beta}_{j,\mathbf{W}}^p$.

From this we get that $r_{j,\mathbf{D}_j^{p,\kappa}}^p$ is minimized for $\ddot{\beta}_j^p = \hat{\beta}_j^{p,\mathrm{opt}}$, $r_{j,\mathbf{D}_j^{p,\hat{\kappa}}}^p$ is minimized for $\ddot{\beta}_j^p = \hat{\beta}_j^p$, and $r_{j,\mathbf{D}_j^p}^p$ is minimized for $\ddot{\beta}_j^p = \dot{\beta}_j^p$. This motivates the use of the three estimators.

Putting $\lambda_j^p = 0$ in (10.8) yields

$$\left[\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p)\right]^{-1} = \mathrm{diag}\left\{\left(\frac{v_{kj\cdot}}{\varphi_j^p}\right)_{\forall k \in \mathcal{A}_j^p}\right\} = \frac{v_{\cdot j\cdot}^p}{\varphi_j^p}\mathbf{D}_j^p$$

and inserting this into (10.7) we get

$$\begin{aligned}\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}} &= \left\{(\mathbf{T}_j^p)'\frac{v_{\cdot j\cdot}^p}{\varphi_j^p}\mathbf{D}_j^p\mathbf{T}_j^p\right\}^{-1}(\mathbf{T}_j^p)'\frac{v_{\cdot j\cdot}^p}{\varphi_j^p}\mathbf{D}_j^p\mathbf{Y}_{\Sigma j}^p \\ &= \left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{Y}_{\Sigma j}^p.\end{aligned}$$

But this is exactly the right hand side of (10.12). Therefore we can conclude that if $\lambda_j^p = 0$ then $\dot{\boldsymbol{\beta}}_j^p = \hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}$, which means that $\dot{\boldsymbol{\beta}}_j^p$ is optimal in this case, and there is reason to hope that $\dot{\boldsymbol{\beta}}_j^p$ is a good estimator if λ_j^p is close to 0. (How good it is depends of course on how stable $\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}$ is around $\lambda_j^p = 0$).

Time-homogeneous model

Analogous to (10.12) we have that an unbiased estimator of $\boldsymbol{\beta}$ based on $\mathbf{Y}_\Sigma$ is given by

$$\dot{\boldsymbol{\beta}} = (\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D}\mathbf{Y}_\Sigma. \tag{10.14}$$

The unbiasedness is a result of

$$\mathrm{E}(\mathbf{Y}_\Sigma) = \mathbf{T}\boldsymbol{\beta}.$$

By using (10.13) we can motivate the use of (10.14) in the same way as we did with (10.12).

An unbiased estimator of $\boldsymbol{\beta}$ based on $\mathbf{Y}_{\Sigma j}$ is given by (10.12), that is,

$$\dot{\boldsymbol{\beta}}_j = \left(\mathbf{T}_j'\mathbf{D}_j\mathbf{T}_j\right)^{-1}\mathbf{T}_j'\mathbf{D}_j\mathbf{Y}_{\Sigma j}. \tag{10.15}$$

It is interesting to study the relationship between (10.14) and (10.15). In fact, we are now going to show the identity

$$\dot{\boldsymbol{\beta}} = \sum_{j=1}^{n}\mathbf{U}_j^{\beta}\dot{\boldsymbol{\beta}}_j \tag{10.16}$$

where

$$\mathbf{U}_j^{\beta,(q\times q)} = (\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\frac{w_{\cdot j\cdot}}{w_{\cdots}}\mathbf{T}_j'\mathbf{D}_j\mathbf{T}_j \quad (j = 1,\ldots,n) \tag{10.17}$$

and

$$\sum_{j=1}^{n}\mathbf{U}_j^{\beta} = \mathbf{I}_q.$$

We begin by calculating

$$\mathbf{T}'\mathbf{D}\mathbf{Y}_\Sigma = \sum_{k=1}^{K}(\mathbf{T}^{(k)})'\mathbf{D}^{(k)}{}_{rk}\mathbf{Y}_{k\Sigma} = \sum_{k=1}^{K}\sum_{j\in\mathcal{B}_k}\mathbf{t}_k\frac{w_{kj\cdot}}{w_{\cdots}}Y_{kj\cdot} \tag{10.18}$$

and

$$\sum_{j=1}^{n}\frac{w_{\cdot j\cdot}}{w_{\cdots}}\mathbf{T}_j'\mathbf{D}_j\mathbf{Y}_{\Sigma j} = \sum_{j=1}^{n}\frac{w_{\cdot j\cdot}}{w_{\cdots}}\sum_{k\in\mathcal{A}_j}\mathbf{t}_k\frac{w_{kj\cdot}}{w_{\cdot j\cdot}}Y_{kj\cdot} = \sum_{j=1}^{n}\sum_{k\in\mathcal{A}_j}\mathbf{t}_k\frac{w_{kj\cdot}}{w_{\cdots}}Y_{kj\cdot}$$

and by changing the order of summation we have

$$\sum_{j=1}^{n}\frac{w_{\cdot j\cdot}}{w_{\cdots}}\mathbf{T}_j'\mathbf{D}_j\mathbf{Y}_{\Sigma j} = \sum_{k=1}^{K}\sum_{j\in\mathcal{B}_k}\mathbf{t}_k\frac{w_{kj\cdot}}{w_{\cdots}}Y_{kj\cdot}. \tag{10.19}$$

Comparing (10.19) with (10.18) we see that

$$\sum_{j=1}^{n} \frac{w_{\cdot j \cdot}}{w_{\dots}} \mathbf{T}'_j \mathbf{D}_j \mathbf{Y}_{\Sigma j} = \mathbf{T}'\mathbf{D}\mathbf{Y}_{\Sigma}. \tag{10.20}$$

Therefore we have from (10.17), (10.15) and (10.14)

$$\begin{aligned}
\sum_{j=1}^{n} \mathbf{U}_j^{\beta} \dot{\boldsymbol{\beta}}_j &= \sum_{j=1}^{n} (\mathbf{T}'\mathbf{D}\mathbf{T})^{-1} \frac{w_{\cdot j \cdot}}{w_{\dots}} \mathbf{T}'_j \mathbf{D}_j \mathbf{T}_j \left(\mathbf{T}'_j \mathbf{D}_j \mathbf{T}_j\right)^{-1} \mathbf{T}'_j \mathbf{D}_j \mathbf{Y}_{\Sigma j} \\
&= (\mathbf{T}'\mathbf{D}\mathbf{T})^{-1} \sum_{j=1}^{n} \frac{w_{\cdot j \cdot}}{w_{\dots}} \mathbf{T}'_j \mathbf{D}_j \mathbf{Y}_{\Sigma j} \overset{(1)}{=} (\mathbf{T}'\mathbf{D}\mathbf{T})^{-1} \mathbf{T}'\mathbf{D}\mathbf{Y}_{\Sigma} \\
&= \dot{\boldsymbol{\beta}}
\end{aligned}$$

where (1) is a consequence of (10.20) but this is exactly (10.16).

Analogous to (10.11) we have that

$$\hat{\boldsymbol{\beta}}_j = \left((\mathbf{T}_j)'\mathbf{D}_j^{\hat{\kappa}}\mathbf{T}_j\right)^{-1} (\mathbf{T}_j)'\mathbf{D}_j^{\hat{\kappa}}\mathbf{Y}_{\Sigma j} \tag{10.21}$$

with

$$\mathbf{D}_j^{\hat{\kappa},(K_j \times K_j)} = \operatorname{diag}\left\{ \left(\frac{v_{kj\cdot}}{v_{kj\cdot} + \hat{\kappa}} \right)_{\forall k \in \mathcal{A}_j} \right\}$$

and $\hat{\kappa} = \hat{\varphi}/\hat{\lambda}$, is an estimator of $\boldsymbol{\beta}$ based on $\mathbf{Y}_{\Sigma j}$.

The optimal (with respect to minimizing the variance) linear unbiased estimator of $\boldsymbol{\beta}$ based on the observations $\mathbf{Y}_{\Sigma}$ is given by

$$\hat{\boldsymbol{\beta}}^{\text{opt}} = \left\{\mathbf{T}' \left[\text{Cov}(\mathbf{Y}_{\Sigma})\right]^{-1} \mathbf{T}\right\}^{-1} \mathbf{T}' \left[\text{Cov}(\mathbf{Y}_{\Sigma})\right]^{-1} \mathbf{Y}_{\Sigma}. \tag{10.22}$$

One could proceed as in the time-heterogeneous models by inserting the expression for $\text{Cov}(\mathbf{Y}_{\Sigma})$ into (10.22) and end up with the analogue to (10.11). The problem with this approach is that the dimension N of $\mathbf{Y}_{\Sigma}$ could be very large in which one in practice would get serious problems when inverting the estimated matrix of $\text{Cov}(\mathbf{Y}_{\Sigma})$. To avoid this problem we will instead take as our starting point the optimal linear unbiased estimators of $\boldsymbol{\beta}$ based on the observations $\mathbf{Y}_{\Sigma j}$ for all the years ($j = 1, \dots, n$) and use as our estimator of $\boldsymbol{\beta}$ a weighted mean of these estimators. Using (10.16) as a motivation, we end up with this estimator of $\boldsymbol{\beta}$ based on $\mathbf{Y}_{\Sigma}$:

$$\hat{\boldsymbol{\beta}} = \sum_{j=1}^{n} \mathbf{U}_j^{\beta} \hat{\boldsymbol{\beta}}_j$$

with $\hat{\boldsymbol{\beta}}_j$ given by (10.21) and $\mathbf{U}_j^{\beta}$ by (10.17).

10.2.3 Estimators of λ_j^p

Time-heterogeneous models

From

$$\lambda_j^p = \text{Var}[m_{kj}(\Theta_{kj})] = \text{E}(m_{kj}(\Theta_{kj}) - \mu_{kj})^2$$

we see that λ_j^p measures how good μ_{kj} is as an estimator of $m_{kj}(\Theta_{kj})$. Further

$$\lambda_j^p = \text{E}[\text{E}(Y_{kji}|\Theta_{kj}) - (\mathbf{t}_{kj}^p)'\boldsymbol{\beta}_j^p]^2$$

so it is natural to base our estimator of λ_j^p $(j = 1, \ldots, n)$ on the class of pseudo-statistics

$$Q_j^{p,\kappa,(1\times 1)} = \left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right)' \mathbf{U}_j^{p,\lambda}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right), \quad (j = 1, \ldots, n) \tag{10.23}$$

where $\mathbf{U}_j^{p,\lambda,(K_j^p\times K_j^p)}$ is a known weighting matrix and $\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}$ is given by (10.10). Thus, we want to calculate

$$\begin{aligned}
\mathrm{E}(Q_j^{p,\kappa}) &\overset{(1)}{=} \mathrm{E}(\mathrm{tr}\,\{Q_j^{p,\kappa}\}) \overset{(2)}{=} \mathrm{E}(\mathrm{tr}\left\{\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right)' \mathbf{U}_j^{p,\lambda}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right)\right\}) \\
&\overset{(3)}{=} \mathrm{E}(\mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right)\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right)'\right\}) \\
&\overset{(4)}{=} \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathrm{E}[\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right)\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}\right)']\right\} \\
&\overset{(5)}{=} \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}})\right\} \\
&\overset{(6)}{=} \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\left[(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right]^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{Y}_{\Sigma j}^p)\right\} \\
&= \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathrm{Cov}(\left\{\mathbf{I}_{K_j^p} - \mathbf{T}_j^p\left[(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right]^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\right\}\mathbf{Y}_{\Sigma j}^p)\right\}.
\end{aligned}$$

In (1) we have used the fact that a scalar equals its trace, in (2) we have inserted (10.23), in (3) we have used (2.1), (4) is a consequence of the fact that the trace operator is linear, in (5) we have utilized that $\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}$ is unbiased and the definition of the covariance operator, and in (6) we have inserted (10.10). Hence we obtain

$$\mathrm{E}(Q_j^{p,\kappa}) = \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^{p,\kappa}\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p)(\mathbf{A}_j^{p,\kappa})'\right\} \tag{10.24}$$

with

$$\mathbf{A}_j^{p,\kappa,(K_j^p\times K_j^p)} = \mathbf{I}_{K_j^p} - \mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}. \tag{10.25}$$

By comparing (10.8) with (10.9) we have

$$\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p) = \lambda_j^p\left(\mathbf{D}_j^{p,\kappa}\right)^{-1}. \tag{10.26}$$

Substituting (10.26) into (10.24) yields

$$\mathrm{E}(Q_j^{p,\kappa}) = \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\lambda_j^p\mathbf{A}_j^{p,\kappa}(\mathbf{D}_j^{p,\kappa})^{-1}(\mathbf{A}_j^{p,\kappa})'\right\}. \tag{10.27}$$

From (10.25) combined with elementary matrix manipulations, we have

$$\begin{aligned}
\mathbf{A}_j^{p,\kappa}(\mathbf{D}_j^{p,\kappa})^{-1}(\mathbf{A}_j^{p,\kappa})' &= \left[\mathbf{I}_{K_j^p} - \mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\right](\mathbf{D}_j^{p,\kappa})^{-1} \\
&\quad\cdot\left[\mathbf{I}_{K_j^p} - \mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right] \\
&= \left[(\mathbf{D}_j^{p,\kappa})^{-1} - \mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right] \\
&\quad\cdot\left[\mathbf{I}_{K_j^p} - \mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right]
\end{aligned}$$

$$
\begin{aligned}
&= (\mathbf{D}_j^{p,\kappa})^{-1} - \mathbf{T}_j^p \left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)' - \mathbf{T}_j^p \left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)' \\
&\quad + \mathbf{T}_j^p \left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p \left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)' \\
&= (\mathbf{D}_j^{p,\kappa})^{-1} - \mathbf{T}_j^p \left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)' \\
&= (\mathbf{D}_j^{p,\kappa})^{-1}\left[\mathbf{I}_{K_j^p} - \mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p \left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\kappa}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right].
\end{aligned}
$$

Substituting (10.25) once again we get

$$\mathbf{A}_j^{p,\kappa}(\mathbf{D}_j^{p,\kappa})^{-1}(\mathbf{A}_j^{p,\kappa})' = (\mathbf{D}_j^{p,\kappa})^{-1}(\mathbf{A}_j^{p,\kappa})' \tag{10.28}$$

Inserting (10.28) into (10.27) we obtain

$$\mathrm{E}(Q_j^{p,\kappa}) = \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\lambda_j^p(\mathbf{D}_j^{p,\kappa})^{-1}(\mathbf{A}_j^{p,\kappa})'\right\}. \tag{10.29}$$

From (10.9) we have

$$\lambda_j^p\left(\mathbf{D}_j^{p,\kappa}\right)^{-1} = \mathrm{diag}\left\{\left(\frac{\varphi_j^p}{v_{kj\cdot}} + \lambda_j^p\right)_{\forall k\in\mathcal{A}_j^p}\right\} = \left(\frac{\varphi_j^p}{v_{\cdot j\cdot}^p}\right)(\mathbf{D}_j^p)^{-1} + \lambda_j^p\mathbf{I}_{K_j^p} \tag{10.30}$$

which inserted into (10.29) yields

$$
\begin{aligned}
\mathrm{E}(Q_j^{p,\kappa}) &= \mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}\left[\left(\frac{\varphi_j^p}{v_{\cdot j\cdot}^p}\right)(\mathbf{D}_j^p)^{-1} + \lambda_j^p\mathbf{I}_{K_j^p}\right](\mathbf{A}_j^{p,\kappa})'\right\} \\
&= \left(\frac{\varphi_j^p}{v_{\cdot j\cdot}^p}\right)\mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^{p,\kappa})'\right\} + \lambda_j^p\mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{A}_j^{p,\kappa})'\right\}.
\end{aligned}
$$

Thus, a class of unbiased pseudo-estimators of λ_j^p $(j = 1, \ldots, n)$ is given by

$$\check{\lambda}_j^{p,\kappa} = \frac{Q_j^{p,\kappa} - (\hat{\varphi}_j^p/v_{\cdot j\cdot}^p)\,\mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^{p,\kappa})'\right\}}{\mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{A}_j^{p,\kappa})'\right\}},$$

where $\hat{\varphi}_j^p$ is given by (10.4). A class of genuine estimators of λ_j^p is therefore given by

$$\check{\lambda}_j^{p,\hat{\kappa}} = \frac{Q_j^{p,\hat{\kappa}} - (\hat{\varphi}_j^p/v_{\cdot j\cdot}^p)\,\mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^{p,\hat{\kappa}})'\right\}}{\mathrm{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{A}_j^{p,\hat{\kappa}})'\right\}}, \tag{10.31}$$

where $Q_j^{p,\hat{\kappa}}$ is given by (10.23) with $\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}$ replaced by $\hat{\boldsymbol{\beta}}_j^p$ given by (10.11), $\hat{\varphi}_j^p$ is given by (10.4). We have defined

$$\mathbf{A}_j^{p,\hat{\kappa},(K_j^p\times K_j^p)} = \mathbf{I}_{K_j^p} - \mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}. \tag{10.32}$$

Since estimator (10.31) may take negative values, whereas $\lambda_j^p \geq 0$ $\forall j$, we prefer the class

$$\hat{\lambda}_j^{p,\hat{\kappa}} = \max(0, \check{\lambda}_j^{p,\hat{\kappa}}). \quad (j = 1, \ldots, n) \tag{10.33}$$

This means that if the observations do not support the hypothesis that $m_{kj}(\Theta_{kj}) \neq \mu_{kj}$ almost surely then we treat the observations as if they indicated that $m_{kj}(\Theta_{kj}) = \mu_{kj}$ almost surely.

Instead of having to take special care of the case $\hat{\lambda}_j^p = 0$ we may replace (10.33) by

$$\hat{\lambda}_j^{p,\hat{\kappa}} = \begin{cases} \check{\lambda}_j^{p,\hat{\kappa}} & \text{if } \check{\lambda}_j^{p,\hat{\kappa}} > 0 \\ \varepsilon / K_j^p & \text{if } \check{\lambda}_j^{p,\hat{\kappa}} \leq 0 \end{cases} \tag{10.34}$$

for $\varepsilon > 0$ "small".

Reasonable weighting matrices may e.g. be

$$\mathbf{U}_j^{p,\lambda} = \mathbf{D}_j^p, \quad \mathbf{U}_j^{p,\lambda} = \mathbf{D}_j^{p,\hat{\kappa}} \quad \text{or} \quad \mathbf{U}_j^{p,\lambda} = \operatorname{diag}\left\{ \left(\frac{I_{kj}}{\sum_{r \in \mathcal{A}_j^p} I_{rj}} \right)_{\forall k \in \mathcal{A}_j^p} \right\}.$$

$\hat{\lambda}_j^{p,\hat{\kappa}}$ given by (10.33) has less expected quadratic loss than $\check{\lambda}_j^{p,\hat{\kappa}}$ given by (10.31). This is shown by considering

$$\begin{aligned} \mathrm{E}(\lambda_j^p - \hat{\lambda}_j^{p,\hat{\kappa}})^2 &= \mathrm{E}\left[(\lambda_j^p - \hat{\lambda}_j^{p,\hat{\kappa}})^2 I(\check{\lambda}_j^{p,\hat{\kappa}} \geq 0)\right] + \mathrm{E}\left[(\lambda_j^p - \hat{\lambda}_j^{p,\hat{\kappa}})^2 I(\check{\lambda}_j^{p,\hat{\kappa}} < 0)\right] \\ &< \mathrm{E}\left[(\lambda_j^p - \check{\lambda}_j^{p,\hat{\kappa}})^2 I(\check{\lambda}_j^{p,\hat{\kappa}} \geq 0)\right] + \mathrm{E}\left[(\lambda_j^p - \check{\lambda}_j^{p,\hat{\kappa}})^2 I(\check{\lambda}_j^{p,\hat{\kappa}} < 0)\right] \end{aligned}$$

where the inequality follows from (10.33). This yields

$$\mathrm{E}(\lambda_j^p - \hat{\lambda}_j^{p,\hat{\kappa}})^2 < \mathrm{E}(\lambda_j^p - \check{\lambda}_j^{p,\hat{\kappa}})^2. \tag{10.35}$$

Since we use expected quadratic loss as optimality criterion this implies that $\hat{\lambda}_j^{p,\hat{\kappa}}$ is a better estimator than $\check{\lambda}_j^{p,\hat{\kappa}}$.

In the sequel alternative estimators of λ_j^p (and of λ in the time-homogeneous model) will be proposed. These estimators can be adjusted in the same way as we did with $\check{\lambda}_j^{p,\hat{\kappa}}$ in (10.33) and (10.34). In addition, (10.35) holds for these estimators as well.

For the specific choice of weighting matrix $\mathbf{U}_j^{p,\lambda} = \mathbf{D}_j^p$ we get

$$\begin{aligned} \operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^{p,\hat{\kappa}})'\right\} &= \operatorname{tr}\left\{\mathbf{D}_j^p(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^{p,\hat{\kappa}})'\right\} = \operatorname{tr}\left\{(\mathbf{A}_j^{p,\hat{\kappa}})'\right\} \\ &\overset{(1)}{=} \operatorname{tr}\left\{\mathbf{I}_{K_j^p} - \mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right\} \\ &\overset{(2)}{=} \operatorname{tr}\left\{\mathbf{I}_{K_j^p}\right\} - \operatorname{tr}\left\{\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right\} \\ &= \operatorname{tr}\left\{\mathbf{I}_{K_j^p}\right\} - \operatorname{tr}\left\{\mathbf{I}_{q_j^p}\right\} \end{aligned}$$

where we in (1) have inserted (10.32), and in (2) we have utilized that the trace operator is linear and (2.1). Since the trace of an identity matrix is its dimension we get

$$\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^{p,\hat{\kappa}})'\right\} = K_j^p - q_j^p. \tag{10.36}$$

Further,

$$\begin{aligned} \operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{A}_j^{p,\hat{\kappa}})'\right\} &= \operatorname{tr}\left\{\mathbf{D}_j^p(\mathbf{A}_j^{p,\hat{\kappa}})'\right\} \\ &\overset{(1)}{=} \operatorname{tr}\left\{\mathbf{D}_j^p\left[\mathbf{I}_{K_j^p} - \mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right]\right\} \\ &= \operatorname{tr}\left\{\mathbf{D}_j^p\right\} - \operatorname{tr}\left\{\mathbf{D}_j^p\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right\} \end{aligned}$$

where we in (1) have substituted (10.32). From the definition of $\mathbf{D}_j^p$ and (2.1) we have

$$\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{A}_j^{p,\hat{\kappa}})'\right\} = 1 - \operatorname{tr}\left\{\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right\}. \tag{10.37}$$

Insertion of (10.36) and (10.37) into (10.31) gives

$$\check{\lambda}_j^{p,\hat{\kappa}} = \frac{Q_j^{p,\hat{\kappa}} - (\hat{\varphi}_j^p/v_{\cdot j\cdot}^p)\left(K_j^p - q_j^p\right)}{1 - \operatorname{tr}\left\{\left((\mathbf{T}_j^p)'\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{D}_j^{p,\hat{\kappa}}\mathbf{T}_j^p\right\}}. \tag{10.38}$$

A disadvantage of (10.31) is that it depends on an estimator of λ_j^p through $\mathbf{D}_j^{p,\hat{\kappa}}$. We may derive an estimator of λ_j^p which does not depend on an estimator of λ_j^p by basing our estimation on

$$Q_j^{p,(1\times1)} = \left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\dot{\boldsymbol{\beta}}_j^p\right)'\mathbf{U}_j^{p,\lambda}\left(\mathbf{Y}_{\Sigma j}^p - \mathbf{T}_j^p\dot{\boldsymbol{\beta}}_j^p\right), \tag{10.39}$$

where $\dot{\boldsymbol{\beta}}_j^p$ is given by (10.12), instead of on (10.23).

By comparing (10.10) with (10.12) we see that we get $\dot{\boldsymbol{\beta}}_j^p$ from $\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}$ by replacing $\mathbf{D}_j^{p,\kappa}$ with $\mathbf{D}_j^p$ in (10.10). The derivations made above can therefore be used to find the expectation of Q_j^p and thereby an alternative estimator of λ_j^p by just replacing $\mathbf{D}_j^{p,\kappa}$ with $\mathbf{D}_j^p$ in $\hat{\boldsymbol{\beta}}_j^{p,\mathrm{opt}}$. From (10.24) we then get

$$\mathrm{E}(Q_j^p) = \operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^p\mathrm{Cov}(\mathbf{Y}_{\Sigma j}^p)(\mathbf{A}_j^p)'\right\} \tag{10.40}$$

where we have defined

$$\mathbf{A}_j^{p,(K_j^p\times K_j^p)} = \mathbf{I}_{K_j^p} - \mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p. \tag{10.41}$$

Inserting (10.26) and (10.30) into (10.40) we get

$$\mathrm{E}(Q_j^p) = \left(\frac{\varphi_j^p}{v_{\cdot j\cdot}^p}\right)\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^p(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^p)'\right\} + \lambda_j^p\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^p(\mathbf{A}_j^p)'\right\}. \tag{10.42}$$

Like (10.28) we have

$$\mathbf{A}_j^p(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^p)' = (\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^p)'$$

and by insertion into (10.42) we obtain

$$\mathrm{E}(Q_j^p) = \left(\frac{\varphi_j^p}{v_{\cdot j\cdot}^p}\right)\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^p)'\right\} + \lambda_j^p\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^p(\mathbf{A}_j^p)'\right\}. \tag{10.43}$$

Thus, a class of unbiased estimators of λ_j^p $(j = 1, \ldots, n)$ is given by

$$\check{\lambda}_j^p = \frac{Q_j^p - (\hat{\varphi}_j^p/v_{\cdot j\cdot}^p)\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^p)'\right\}}{\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^p(\mathbf{A}_j^p)'\right\}}, \quad (j = 1, \ldots, n) \tag{10.44}$$

where $\hat{\varphi}_j^p$ is given by (10.4).

From (10.36) we see that for the specific choice of weighting matrix $\mathbf{U}_j^{p,\lambda} = \mathbf{D}_j^p$ we have

$$\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}(\mathbf{D}_j^p)^{-1}(\mathbf{A}_j^p)'\right\} = K_j^p - q_j^p. \tag{10.45}$$

Further, we have

$$\begin{aligned}
\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^p(\mathbf{A}_j^p)'\right\} &= \operatorname{tr}\left\{\mathbf{D}_j^p\mathbf{A}_j^p(\mathbf{A}_j^p)'\right\} \\
&= \operatorname{tr}\left\{\mathbf{D}_j^p\left[\mathbf{I}_{K_j^p} - \mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\right]\right. \\
&\quad \left.\cdot\left[\mathbf{I}_{K_j^p} - \mathbf{D}_j^p\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right]\right\} \\
&= \operatorname{tr}\left\{\mathbf{D}_j^p\right\} - \operatorname{tr}\left\{\mathbf{D}_j^p\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\right\} \\
&\quad - \operatorname{tr}\left\{\mathbf{D}_j^p\mathbf{D}_j^p\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right\} \\
&\quad + \operatorname{tr}\left\{\mathbf{D}_j^p\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{D}_j^p\mathbf{T}_j^p\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\right\} \\
&= \operatorname{tr}\left\{\mathbf{D}_j^p\right\} - \operatorname{tr}\left\{\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{D}_j^p\mathbf{T}_j^p\right\}
\end{aligned}$$

and we obtain

$$\operatorname{tr}\left\{\mathbf{U}_j^{p,\lambda}\mathbf{A}_j^p(\mathbf{A}_j^p)'\right\} = 1 - \operatorname{tr}\left\{\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'(\mathbf{D}_j^p)^2\mathbf{T}_j^p\right\}. \tag{10.46}$$

We insert (10.45) and (10.46) into (10.44) and get

$$\check{\lambda}_j^p = \frac{Q_j^p - (\hat{\varphi}_j^p/v_{\cdot j\cdot}^p)\left(K_j^p - q_j^p\right)}{1 - \operatorname{tr}\left\{\left((\mathbf{T}_j^p)'\mathbf{D}_j^p\mathbf{T}_j^p\right)^{-1}(\mathbf{T}_j^p)'(\mathbf{D}_j^p)^2\mathbf{T}_j^p\right\}}. \quad (j = 1, \ldots, n) \tag{10.47}$$

Compare (10.47) with (10.38) and notice the analogy between them.

Time-homogeneous model

We base our estimation of λ on the class of statistics

$$Q^{(1)} = \left(\mathbf{Y}_\Sigma - \mathbf{T}\dot{\boldsymbol{\beta}}\right)'\mathbf{U}^\lambda\left(\mathbf{Y}_\Sigma - \mathbf{T}\dot{\boldsymbol{\beta}}\right) \tag{10.48}$$

where $\dot{\boldsymbol{\beta}}$ is given by (10.14) and $\mathbf{U}^{\lambda,(N\times N)}$ is a known weighting matrix. By analogy, (10.44) gives us the estimator

$$\check{\lambda}^{\hat{\varrho}} = \frac{Q^{(1)} - (\hat{\varphi}/v_{\cdots})\operatorname{tr}\left\{\mathbf{U}^\lambda\mathbf{D}^{-1}\mathbf{B}'\right\}}{\operatorname{tr}\left\{\mathbf{U}^\lambda\mathbf{B}\mathbf{A}_{\hat{\varrho}}\mathbf{B}'\right\}} \tag{10.49}$$

with $\mathbf{B}$ and $\mathbf{A}_{\hat{\varrho}}$ given by

$$\begin{aligned}
\mathbf{B}^{(N\times N)} &= \mathbf{I}_N - \mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D} && (10.50)\\
\mathbf{A}_{\hat{\varrho}}^{(N\times N)} &= \operatorname{diag}\left\{\left(\mathbf{A}_{\hat{\varrho}}^{(1)}, \ldots, \mathbf{A}_{\hat{\varrho}}^{(K)}\right)'\right\} && (10.51)\\
\mathbf{A}_{\hat{\varrho}}^{(k),(n_k\times n_k)} &= \begin{pmatrix} 1 & \hat{\varrho} & \ldots & \hat{\varrho}^{n_k-1} \\ \hat{\varrho} & 1 & \ldots & \hat{\varrho}^{n_k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varrho}^{n_k-1} & \hat{\varrho}^{n_k-2} & \ldots & 1 \end{pmatrix}. \quad (k = 1, \ldots, K) && (10.52)
\end{aligned}$$

$Q^{(1)}$ is given by (10.48), $\hat{\varphi}$ is given by (10.6), and $\hat{\varrho}$ is some estimator of ϱ.

From (10.51) and (10.52) we see that if $\hat{\varrho} = 0$ then $\mathbf{A}_{\hat{\varrho}} = \mathbf{I}_N$ and (10.49) becomes

$$\check{\lambda}^{\hat{\varrho}} = \frac{Q^{(1)} - (\hat{\varphi}/v_{\cdots})\,\mathrm{tr}\left\{\mathbf{U}^{\lambda}\mathbf{D}^{-1}\mathbf{B}'\right\}}{\mathrm{tr}\left\{\mathbf{U}^{\lambda}\mathbf{B}\mathbf{B}'\right\}}.$$

This estimator is the complete analogue to estimator (10.44). This is reasonable since $\varrho = 0$ means that observations from different years are (unconditionally) uncorrelated and hence there is a symmetry between observations from different years and observations from different car models (from Assumption 4.1 we have that observations from different car models are uncorrelated).

For the specific choice of weighting matrix $\mathbf{U}^{\lambda} = \mathbf{D}$ we have from (10.45), by analogy,

$$\mathrm{tr}\left\{\mathbf{U}^{\lambda}\mathbf{D}^{-1}\mathbf{B}'\right\} = N - q. \tag{10.53}$$

Further, we have from (10.50)

$$\begin{aligned}
\mathbf{B}'\mathbf{D}\mathbf{B} &= [\mathbf{I}_N - \mathbf{D}\mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}']\mathbf{D}[\mathbf{I}_N - \mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D}] \\
&= \mathbf{D} - \mathbf{D}\mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D} - \mathbf{D}\mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D} \\
&\quad + \mathbf{D}\mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D}\mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D} \\
&= \mathbf{D} - \mathbf{D}\mathbf{T}(\mathbf{T}'\mathbf{D}\mathbf{T})^{-1}\mathbf{T}'\mathbf{D}
\end{aligned}$$

and using (10.50) once again we get

$$\mathbf{B}'\mathbf{D}\mathbf{B} = \mathbf{D}\mathbf{B}. \tag{10.54}$$

Using (2.1) we get

$$\mathrm{tr}\left\{\mathbf{U}^{\lambda}\mathbf{B}\mathbf{A}_{\hat{\varrho}}\mathbf{B}'\right\} = \mathrm{tr}\left\{\mathbf{D}\mathbf{B}\mathbf{A}_{\hat{\varrho}}\mathbf{B}'\right\} = \mathrm{tr}\left\{\mathbf{A}_{\hat{\varrho}}\mathbf{B}'\mathbf{D}\mathbf{B}\right\}.$$

Inserting (10.54) we obtain

$$\mathrm{tr}\left\{\mathbf{U}^{\lambda}\mathbf{B}\mathbf{A}_{\hat{\varrho}}\mathbf{B}'\right\} = \mathrm{tr}\left\{\mathbf{A}_{\hat{\varrho}}\mathbf{D}\mathbf{B}\right\}. \tag{10.55}$$

Inserting (10.53) and (10.55) into (10.49) yields

$$\check{\lambda}^{\hat{\varrho}} = \frac{Q^{(1)} - (\hat{\varphi}/v_{\cdots})\,(N-q)}{\mathrm{tr}\left\{\mathbf{A}_{\hat{\varrho}}\mathbf{D}\mathbf{B}\right\}}$$

(10.49) has the deficiency that it is dependent on $\hat{\varrho}$ (this is really a deficiency since, as we shall see in subsection 10.2.4, our estimator of ϱ depends on an estimator of λ). We will therefore try to derive a class of estimators of λ which does not depend on any estimator of ϱ. We begin by considering

$$Q^{(2)} = \sum_{j=1}^{n} u_j^Q Q_j \tag{10.56}$$

with Q_j given by (10.39) and u_j^Q $(j = 1, \ldots, n)$ are reasonable weights. This yields

$$\mathrm{E}(Q^{(2)}) = \sum_{j=1}^{n} u_j^Q \mathrm{E}(Q_j) = \sum_{j=1}^{n} u_j^Q \left\{ \left(\frac{\varphi}{v_{\cdot j \cdot}}\right) \mathrm{tr}\left\{\mathbf{U}_j^{\lambda}\mathbf{D}_j^{-1}\mathbf{A}_j'\right\} + \lambda \mathrm{tr}\left\{\mathbf{U}_j^{\lambda}\mathbf{A}_j\mathbf{A}_j'\right\}\right\}$$

where the latter equality is a consequence of (10.43). $\mathbf{A}_j$ is given by (10.41). Thus we have

$$\mathrm{E}(Q^{(2)}) = \varphi \sum_{j=1}^{n} \frac{u_j^Q}{v_{\cdot j \cdot}} \mathrm{tr}\left\{\mathbf{U}_j^\lambda \mathbf{D}_j^{-1} \mathbf{A}_j'\right\} + \lambda \sum_{j=1}^{n} u_j^Q \mathrm{tr}\left\{\mathbf{U}_j^\lambda \mathbf{A}_j \mathbf{A}_j'\right\}. \tag{10.57}$$

A class of estimators of λ which do not depend on any estimator of λ or ϱ is given by

$$\check{\lambda} = \frac{Q^{(2)} - \hat{\varphi} \sum_{j=1}^{n} (u_j^Q / v_{\cdot j \cdot}) \mathrm{tr}\left\{\mathbf{U}_j^\lambda \mathbf{D}_j^{-1} \mathbf{A}_j'\right\}}{\sum_{j=1}^{n} u_j^Q \mathrm{tr}\left\{\mathbf{U}_j^\lambda \mathbf{A}_j \mathbf{A}_j'\right\}}. \tag{10.58}$$

$Q^{(2)}$, $\hat{\varphi}$ and $\mathbf{A}_j$ are given by (10.56), (10.6) and (10.41), respectively.

From (10.45) and (10.46) we see that for the specific choice $u_j^Q = v_{\cdot j \cdot}/v_{\cdots}$ and $\mathbf{U}_j^\lambda = \mathbf{D}_j$ (10.58) becomes

$$\check{\lambda} = \frac{Q^{(2)} - (\hat{\varphi}/v_{\cdots}) \sum_{j=1}^{n} (K_j - q)}{\sum_{j=1}^{n} (v_{\cdot j \cdot}/v_{\cdots}) \left(1 - \mathrm{tr}\left\{\left(\mathbf{T}_j' \mathbf{D}_j \mathbf{T}_j\right)^{-1} \mathbf{T}_j' \mathbf{D}_j^2 \mathbf{T}_j\right\}\right)}$$

which is equivalent to

$$\check{\lambda} = \frac{Q^{(2)} - (\hat{\varphi}/v_{\cdots})(N - nq)}{1 - \sum_{j=1}^{n} (v_{\cdot j \cdot}/v_{\cdots}) \mathrm{tr}\left\{\left(\mathbf{T}_j' \mathbf{D}_j \mathbf{T}_j\right)^{-1} \mathbf{T}_j' \mathbf{D}_j^2 \mathbf{T}_j\right\}}.$$

We see that (10.58) does not depend on any unknown quantity, only on $\hat{\varphi}$ and other observables. From (10.57) and (10.58) we have that since $\hat{\varphi}$ is an unbiased estimator, so is $\check{\lambda}$.

10.2.4 Estimators of ϱ_j^p

Suppose that we in practice are not willing to assume that Assumption 7.6 are fulfilled for all j. After all, this assumption is primarily made to be able to find a recursive procedure in order to calculate the credibility estimators and to secure that the correlations between years decrease as the distances between the years increase (cf. (5.4)), and not because we feel that it is exactly true in practice. Then we consider Assumption 7.6 rather as a working hypothesis which we do not want to rest upon more than necessary. Thus we will, for the estimation purpose, consider Assumption 7.6 to be true for the latest $d+1$ years, say, only. We could also argue that we, for various reasons, just want to base the estimation of ϱ_i^p ($p = A, B, BA$) on observations from the latest $d+1$ years. We shall see that the implication on our estimator in the two ways of reasoning is the same.

This does not involve any loss of generality. If we want to base the estimation on observations from all the previous years we may put $d = n - 1$.

For notational convenience we define $m(i,d) = i - \min(i,d) + 1$, that is, $m(i,d) = \max(i - d + 1, 1)$.

Time-heterogeneous models

We will now consider estimation of the parameters ϱ_i^A, ϱ_i^{BA} and ϱ_i^B. The idea is to find an unbiased estimator of the covariance (λ_{ij}^p) and from this derive an estimator of ϱ_i^p.

A. Estimators of ϱ_i^A. From Assumption 7.6 we have

$$\lambda_{i+1,j}^A = \varrho_i^A \lambda_{ij}^A. \quad (m(i,d) \le j \le i \;;\; i = 1,\ldots,n-1)$$

Summation over j gives

$$\sum_{j=m(i,d)}^{i} \lambda_{i+1,j}^A = \sum_{j=m(i,d)}^{i} \varrho_i^A \lambda_{ij}^A. \quad (i = 1,\ldots,n-1)$$

Since ϱ_i^A does not depend on the summation index j we have

$$\varrho_i^A = \frac{\sum_{j=m(i,d)}^{i} \lambda_{i+1,j}^A}{\sum_{j=m(i,d)}^{i} \lambda_{ij}^A} = \frac{\sum_{j=m(i,d)}^{i} \mathrm{Cov}(m_{k,i+1}(\Theta_{k,i+1}), m_{kj}(\Theta_{kj}))}{\sum_{j=m(i,d)}^{i-1} \mathrm{Cov}(m_{k'i}(\Theta_{k'i}), m_{k'j}(\Theta_{k'j})) + \lambda_i^A} \tag{10.59}$$

where $k \in \mathcal{A}_{j,i+1}^A$, $k' \in \mathcal{A}_{ij}^A$ and $i = 1,\ldots,n-1$. It is therefore natural to base our estimator of ϱ_i^A $(i = 1,\ldots,n-1)$ on the class of statistics

$$Q_{ij}^A = \left({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A - {}^{i}_{j}\mathbf{T}_i^A \dot{\boldsymbol{\beta}}_i^A\right)' \mathbf{U}_{ij}^{\varrho,A} \left({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A - {}^{i}_{j}\mathbf{T}_j^A \dot{\boldsymbol{\beta}}_j^A\right) \quad (i > j) \tag{10.60}$$

where $\mathbf{U}_{ij}^{\varrho,A,({}^{i}_{j}K^A \times {}^{i}_{j}K^A)}$ is a known weighting matrix and $\dot{\boldsymbol{\beta}}_j^A$ $(j = 1,\ldots,i+1)$ is given by (10.12). We begin by calculating

$$\begin{aligned}
\mathrm{E}(Q_{ij}^A) &= \mathrm{tr}\left\{\mathrm{E}[\left({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A - {}^{i}_{j}\mathbf{T}_i^A \dot{\boldsymbol{\beta}}_i^A\right)' \mathbf{U}_{ij}^{\varrho,A} \left({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A - {}^{i}_{j}\mathbf{T}_j^A \dot{\boldsymbol{\beta}}_j^A\right)]\right\} \\
&= \mathrm{tr}\left\{\mathbf{U}_{ij}^{\varrho,A} \mathrm{Cov}({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A - {}^{i}_{j}\mathbf{T}_j^A \dot{\boldsymbol{\beta}}_j^A, \left({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A - {}^{i}_{j}\mathbf{T}_i^A \dot{\boldsymbol{\beta}}_i^A\right)')\right\} \\
&= \mathrm{tr}\left\{\mathbf{U}_{ij}^{\varrho,A} \mathrm{Cov}({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A - {}^{i}_{j}\mathbf{A}_j^A \mathbf{Y}_{\Sigma j}^A, \left({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A - {}^{i}_{j}\mathbf{A}_i^A \mathbf{Y}_{\Sigma i}^A\right)')\right\}
\end{aligned}$$

with

$${}^{i}_{j}\mathbf{A}_h^{A,({}^{i}_{j}K^A \times K_h^A)} = {}^{i}_{j}\mathbf{T}_h^A \left((\mathbf{T}_h^A)'\mathbf{D}_h^A \mathbf{T}_h^A\right)^{-1} (\mathbf{T}_h^A)'\mathbf{D}_h^A. \quad (h = i,j)$$

This gives

$$\begin{aligned}
\mathrm{E}(Q_{ij}^A) = {} & \mathrm{tr}\left\{\mathbf{U}_{ij}^{\varrho,A}\left[\mathrm{Cov}\left({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A, ({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A)'\right) - \mathrm{Cov}\left({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A, (\mathbf{Y}_{\Sigma i}^A)'\right)({}^{i}_{j}\mathbf{A}_i^A)'\right]\right\} \\
& - \mathrm{tr}\left\{\mathbf{U}_{ij}^{\varrho,A}\left[{}^{i}_{j}\mathbf{A}_j^A \mathrm{Cov}\left(\mathbf{Y}_{\Sigma j}^A, ({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A)'\right) - {}^{i}_{j}\mathbf{A}_j^A \mathrm{Cov}\left(\mathbf{Y}_{\Sigma j}^A, (\mathbf{Y}_{\Sigma i}^A)'\right)({}^{i}_{j}\mathbf{A}_i^A)'\right]\right\}.
\end{aligned}$$

We have from (A.9) $(i > j)$

$$\begin{aligned}
\mathrm{Cov}\left({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A, ({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A)'\right)^{({}^{i}_{j}K^A \times {}^{i}_{j}K^A)} &= \lambda_{ij}^A \mathbf{I}_{{}^{i}_{j}K^A} \\
\mathrm{Cov}\left({}^{i}_{j}\mathbf{Y}_{\Sigma j}^A, (\mathbf{Y}_{\Sigma i}^A)'\right)^{({}^{i}_{j}K^A \times K_i^A)} &= \lambda_{ij}^A \boldsymbol{\Delta}_{ij}^{(1),A} \quad \text{where } \boldsymbol{\Delta}_{ij}^{(1),A,({}^{i}_{j}K^A \times K_i^A)} = (\delta_{k,l})_{\forall k \in \mathcal{A}_{ij}^A}^{\forall l \in \mathcal{A}_i^A} \\
\mathrm{Cov}\left(\mathbf{Y}_{\Sigma j}^A, ({}^{i}_{j}\mathbf{Y}_{\Sigma i}^A)'\right)^{(K_j^A \times {}^{i}_{j}K^A)} &= \lambda_{ij}^A \boldsymbol{\Delta}_{ij}^{(2),A} \quad \text{where } \boldsymbol{\Delta}_{ij}^{(2),A,(K_j^A \times {}^{i}_{j}K^A)} = (\delta_{k,l})_{\forall k \in \mathcal{A}_j^A}^{\forall l \in \mathcal{A}_{ij}^A} \\
\mathrm{Cov}\left(\mathbf{Y}_{\Sigma j}^A, (\mathbf{Y}_{\Sigma i}^A)'\right)^{(K_j^A \times K_i^A)} &= \lambda_{ij}^A \boldsymbol{\Delta}_{ij}^{(3),A} \quad \text{where } \boldsymbol{\Delta}_{ij}^{(3),A,(K_j^A \times K_i^A)} = (\delta_{k,l})_{\forall k \in \mathcal{A}_j^A}^{\forall l \in \mathcal{A}_i^A}
\end{aligned}$$

giving

$$\mathrm{E}(Q_{ij}^A) = \lambda_{ij}^A a_{ij}^A \quad (i > j) \tag{10.61}$$

where

$$a_{ij}^A = 1 - \operatorname{tr}\left\{\mathbf{U}_{ij}^{\varrho,A}\boldsymbol{\Delta}_{ij}^{(1),A}({}_j^i\mathbf{A}_i^A)'\right\} - \operatorname{tr}\left\{\mathbf{U}_{ij}^{\varrho,A}\,{}_j^i\mathbf{A}_j^A\boldsymbol{\Delta}_{ij}^{(2),A}\right\} + \operatorname{tr}\left\{\mathbf{U}_{ij}^{\varrho,A}\,{}_j^i\mathbf{A}_j^A\boldsymbol{\Delta}_{ij}^{(3),A}({}_j^i\mathbf{A}_i^A)'\right\}.$$

From (10.61) we have that

$$\hat{\lambda}_{ij}^A = Q_{ij}^A / a_{ij}^A \quad (i > j) \tag{10.62}$$

is a class of unbiased estimators of λ_{ij}^A $(i > j)$, and the estimation of $\lambda_{i+1,j}^A$ $(m(i,d) \leq j \leq i\ ;\ i = 1, \ldots, n-1)$ is given by the following algorithm.

Algorithm 10.1
For $i = 1, \ldots, n-1$ *do*
Decide upon an estimator of λ_i^A *and compute* $\hat{\lambda}_{ii}^A = \hat{\lambda}_i^A$ *thereof.*
For $j = i, \ldots, m(i,d)$ *do*
Compute $\hat{\lambda}_{i+1,j}^A$ *by* (10.62)*;*

From (10.59) we have that

$${}_d\dot{\varrho}_i^A = \frac{\sum_{j=m(i,d)}^{i} \hat{\lambda}_{i+1,j}^A}{\sum_{j=m(i,d)}^{i} \hat{\lambda}_{ij}^A}, \quad (i = 1, \ldots, n-1) \tag{10.63}$$

with $\hat{\lambda}_{rj}^A$ $(r = i, i+1)$ given by Algorithm 10.1, is a reasonable estimator of ϱ_j^A. From (A.14) we have

$$\pi_{i+1,i}^A = \varrho_i^A \sqrt{\frac{\lambda_i^A}{\lambda_{i+1}^A}}$$

which gives the following reasonable estimator of $\pi_{i+1,i}^A$:

$${}_d\dot{\pi}_{i+1,i}^A = {}_d\dot{\varrho}_i^A \sqrt{\frac{\hat{\lambda}_i^A}{\hat{\lambda}_{i+1}^A}},$$

with ${}_d\dot{\varrho}_i^A$ given by (10.63). Since $\pi_{i+1,i}^A \in [0,1]$ whereas ${}_d\dot{\pi}_{i+1,i}^A$ may attain a value beyond this interval, we prefer the estimator

$${}_d\hat{\pi}_{i+1,i}^A = \begin{cases} 1 & \text{if } {}_d\dot{\pi}_{i+1,i}^A \geq 1 \\ {}_d\dot{\pi}_{i+1,i}^A & \text{if } 0 < {}_d\dot{\pi}_{i+1,i}^A < 1 \\ 0 & \text{if } {}_d\dot{\pi}_{i+1,i}^A \leq 0 \end{cases} \tag{10.64}$$

of $\pi_{i+1,i}^A$. Finally, we estimate ϱ_i^A by

$${}_d\hat{\varrho}_i^A = {}_d\hat{\pi}_{i+1,i}^A \sqrt{\frac{\hat{\lambda}_{i+1}^A}{\hat{\lambda}_i^A}} \quad (i = 1, \ldots, n-1) \tag{10.65}$$

with ${}_d\hat{\pi}_{i+1,i}^A$ given by (10.64).

(5.2) will give us this estimator of $\pi_{i+1,j}^A$ $(j = 1, \ldots, i)$

$${}_d\hat{\pi}_{i+1,j}^A = \prod_{k=j}^{i} {}_d\hat{\pi}_{k+1,k}^A \quad (j = 1, \ldots, i) \tag{10.66}$$

with ${}_d\hat{\pi}^A_{k+1,k}$ given by (10.64).

As weighting matrices in (10.60) we make the following suggestions

$$\mathbf{U}^{\varrho,A}_{ij} = \frac{1}{{}^i_jK^A}\mathbf{I}_{{}^i_jK^A} \qquad \mathbf{U}^{\varrho,A}_{ij} = \operatorname{diag}\left\{\left(\frac{I_{ki}+I_{kj}}{\sum_{r\in\mathcal{A}^A_{ij}}(I_{ri}+I_{rj})}\right)_{\forall k\in\mathcal{A}^A_{ij}}\right\}$$

$$\mathbf{U}^{\varrho,A}_{ij} = \operatorname{diag}\left\{\left(\frac{v_{ki\cdot}+v_{kj\cdot}}{\sum_{r\in\mathcal{A}^A_{ij}}(v_{ri\cdot}+v_{rj\cdot})}\right)_{\forall k\in\mathcal{A}^A_{ij}}\right\}.$$

B. Estimators of ϱ^{BA}_i. We now define $(i>j)$

$$\begin{aligned}
{}_{(i-1)}\mathcal{A}^{BA}_{ij} &= \mathcal{A}^B_{i,i-1}\cap\mathcal{A}^A_j \\
{}^i_jK^{BA}_{i-1} &= \text{number of elements in } {}_{(i-1)}\mathcal{A}^{BA}_{ij} \\
{}^i_j\mathbf{Y}^{BA}_{\Sigma l} &= (Y_{kl\cdot})_{\forall k\in{}_{(i-1)}\mathcal{A}^{BA}_{ij}} \quad (l=j,i) \\
{}^i_j\mathbf{T}^{BA}_l &= (\mathbf{t}_{kl})_{\forall k\in{}_{(i-1)}\mathcal{A}^{BA}_{ij}} \quad (l=j,i) \\
{}^i_j\mathbf{A}^{BA,({}^i_jK^{BA}_{i-1}\times K^p_h)}_h &= {}^i_j\mathbf{T}^{BA}_h\left((\mathbf{T}^p_h)'\mathbf{D}^p_h\mathbf{T}^p_h\right)^{-1}(\mathbf{T}^p_h)'\mathbf{D}^p_h \quad ((h,p)=(i,B),(j,A)) \\
\boldsymbol{\Delta}^{(1),BA,({}^i_jK^{BA}_{i-1}\times K^B_i)}_{ij} &= (\delta_{k,l})^{\forall l\in\mathcal{A}^B_i}_{\forall k\in{}_{(i-1)}\mathcal{A}^{BA}_{ij}} \\
\boldsymbol{\Delta}^{(2),BA,(K^A_j\times{}^i_jK^{BA}_{i-1})}_{ij} &= (\delta_{k,l})^{\forall l\in{}_{(i-1)}\mathcal{A}^{BA}_{ij}}_{\forall k\in\mathcal{A}^A_j} \\
\boldsymbol{\Delta}^{(3),BA,(K^A_j\times K^B_i)}_{ij} &= (\delta_{k,l})^{\forall l\in\mathcal{A}^B_i}_{\forall k\in\mathcal{A}^A_j} \\
Q^{BA}_{ij} &= \left({}^i_j\mathbf{Y}^{BA}_{\Sigma i}-{}^i_j\mathbf{T}^{BA}_i\dot{\boldsymbol{\beta}}^B_i\right)'\mathbf{U}^{\varrho,BA}_{ij}\left({}^i_j\mathbf{Y}^{BA}_{\Sigma j}-{}^i_j\mathbf{T}^{BA}_j\dot{\boldsymbol{\beta}}^A_j\right) \\
a^{BA}_{ij} &= 1-\operatorname{tr}\left\{\mathbf{U}^{\varrho,BA}_{ij}\boldsymbol{\Delta}^{(1),BA}_{ij}({}^i_j\mathbf{A}^{BA}_i)'\right\}-\operatorname{tr}\left\{\mathbf{U}^{\varrho,BA}_{ij}\,{}^i_j\mathbf{A}^{BA}_j\boldsymbol{\Delta}^{(2),BA}_{ij}\right\} \\
&\quad+\operatorname{tr}\left\{\mathbf{U}^{\varrho,BA}_{ij}\,{}^i_j\mathbf{A}^{BA}_j\boldsymbol{\Delta}^{(3),BA}_{ij}({}^i_j\mathbf{A}^{BA}_i)'\right\}
\end{aligned}$$

where $\mathbf{U}^{\varrho,BA,({}^i_jK^{BA}_{i-1}\times{}^i_jK^{BA}_{i-1})}_{ij}$ are known weighting matrices and $\dot{\boldsymbol{\beta}}^p_j$ $(p=A,B\ ;\ j=2,\ldots,n)$ are given by (10.12). From Assumption 7.6 we have

$$\lambda^{(i),BA}_{i+1,j}=\varrho^{BA}_i\lambda^A_{ij} \quad (m(i,d)\le j\le i\ ;\ i=1,\ldots,n-1)$$

implying

$$\varrho^{BA}_i=\frac{\sum_{j=m(i,d)}^i\lambda^{(i),BA}_{i+1,j}}{\sum_{j=m(i,d)}^i\lambda^A_{ij}}. \tag{10.67}$$

Completely analogous to what we did in subsection A we may show that

$$\hat{\lambda}^{(i),BA}_{ij}=Q^{BA}_{ij}/a^{BA}_{ij} \quad (i>j) \tag{10.68}$$

is a class of unbiased estimators of $\lambda^{(i),BA}_{ij}$ $(i>j)$.

The estimation of $\lambda^{(i),BA}_{i+1,j}$ $(m(i,d)\le j\le i\ ;\ i=1,\ldots,n-1)$ is given by the following algorithm.

Algorithm 10.2
For $i = 1, \ldots, n-1$ *do*
Decide upon estimators of λ_i^A *and* λ_{i+1}^B*, and compute* $\hat{\lambda}_{ii}^A = \hat{\lambda}_i^A$ *and* $\hat{\lambda}_{i+1,i+1}^B = \hat{\lambda}_{i+1}^B$ *thereof.*
For $j = i, \ldots, m(i,d)$ *do*
Compute $\hat{\lambda}_{i+1,j}^{(i),BA}$ *by* (10.68)*;*

From (10.67) we have that

$$
{}_d\dot{\varrho}_i^{BA} = \frac{\sum_{j=m(i,d)}^{i} \hat{\lambda}_{i+1,j}^{(i),BA}}{\sum_{j=m(i,d)}^{i} \hat{\lambda}_{ij}^{A}}, \quad (i = 1, \ldots, n-1) \tag{10.69}
$$

with $\hat{\lambda}_{rj}^A$ $(r = i, i+1)$ given by Algorithm 10.1, is a reasonable estimator of ϱ_j^{BA}. From (A.14) we have

$$
\pi_{i+1,i}^{BA} = \varrho_i^{BA} \sqrt{\frac{\lambda_i^A}{\lambda_{i+1}^B}}
$$

which gives the following reasonable estimator of $\pi_{i+1,i}^{BA}$:

$$
{}_d\dot{\pi}_{i+1,i}^{BA} = {}_d\dot{\varrho}_i^{BA} \sqrt{\frac{\hat{\lambda}_i^A}{\hat{\lambda}_{i+1}^B}},
$$

with ${}_d\dot{\varrho}_i^{BA}$ given by (10.69). Since $\pi_{i+1,i}^{BA} \in [0,1]$ whereas ${}_d\dot{\pi}_{i+1,i}^{BA}$ may attain a value beyond this interval, we prefer the estimator

$$
{}_d\hat{\pi}_{i+1,i}^{BA} = \begin{cases} 1 & \text{if } {}_d\dot{\pi}_{i+1,i}^{BA} \geq 1 \\ {}_d\dot{\pi}_{i+1,i}^{BA} & \text{if } 0 < {}_d\dot{\pi}_{i+1,i}^{BA} < 1 \\ 0 & \text{if } {}_d\dot{\pi}_{i+1,i}^{BA} \leq 0 \end{cases} \tag{10.70}
$$

of $\pi_{i+1,i}^{BA}$. Finally, we estimate ϱ_i^{BA} by

$$
{}_d\hat{\varrho}_i^{BA} = {}_d\hat{\pi}_{i+1,i}^{BA} \sqrt{\frac{\hat{\lambda}_{i+1}^B}{\hat{\lambda}_i^A}} \quad (i = 1, \ldots, n-1)
$$

with ${}_d\hat{\pi}_{i+1,i}^{A}$ given by (10.70).

(5.2) will give us this estimator of $\pi_{i+1,j}^{BA}$ $(j = 1, \ldots, i-1)$

$$
{}_d\hat{\pi}_{i+1,j}^{BA} = \hat{\pi}_{i+1,i}^{BA} \prod_{k=j}^{i-1} {}_d\hat{\pi}_{k+1,k}^{A} \quad (j = 1, \ldots, i-1)
$$

with $\hat{\pi}_{i+1,i}^{BA}$ given by (10.70) and ${}_d\hat{\pi}_{k+1,k}^{A}$ given by (10.64).

Reasonable weighting matrices in Q_{ij}^{BA} may e.g. be

$$
\begin{aligned}
\mathbf{U}_{ij}^{\varrho,BA} &= \frac{1}{{}_j^iK_j^{BA}} \mathbf{I}_{{}_j^iK_j^{BA}} \\
\mathbf{U}_{ij}^{\varrho,BA} &= \operatorname{diag}\left\{ \left(\frac{I_{k'i} + I_{k'j}}{\sum_{r \in {}_{(i-1)}\mathcal{A}_{ij}^{BA}} (I_{ri} + I_{rj})} \right)_{\forall k' \in {}_{(i-1)}\mathcal{A}_{ij}^{BA}} \right\} \\
\mathbf{U}_{ij}^{\varrho,BA} &= \operatorname{diag}\left\{ \left(\frac{v_{k'i\cdot} + v_{k'j\cdot}}{\sum_{r \in {}_{(i-1)}\mathcal{A}_{ij}^{BA}} (v_{ri\cdot} + v_{rj\cdot})} \right)_{\forall k' \in {}_{(i-1)}\mathcal{A}_{ij}^{BA}} \right\}.
\end{aligned}
$$

C. Estimators of ϱ_i^B. From Assumption 7.6 we have

$$\lambda_{i+1,j}^B = \varrho_i^B \lambda_{ij}^B \quad (m(i,d) \le j \le i \,;\, i = 1,\ldots,n-1) \tag{10.71}$$

$$\lambda_{i+1,j}^{(r_k),BA} = \varrho_i^B \lambda_{ij}^{(r_k),BA}. \quad (k \in \mathcal{A}_{i+1,i}^B \cap \mathcal{A}_j^A \,;\, m(i,d) \le j < i \,;\, i = 2,\ldots,n-1) \tag{10.72}$$

We base the estimation of ϱ_i^B on (10.71) and proceed in a completely analogous way as in subsection A by replacing A by B in all expressions. Then we get this estimator of ϱ_i^B:

$$_d\hat{\varrho}_i^B = {_d\hat{\pi}_{i+1,i}^B} \sqrt{\frac{\hat{\lambda}_{i+1}^B}{\hat{\lambda}_i^B}} \quad (i = 1,\ldots,n-1)$$

with $_d\hat{\pi}_{i+1,i}^B$ given by (10.64) with A replaced by B.

We may base the estimation on (10.72) as well. This will not be done in this paper. The author feels that deriving an estimator based on (10.72) is not worth the effort. We would have to introduce even more notation than we already have done. In addition, one may question whether (10.72) contributes significantly more information in addition to (10.71); we base the estimation of ϱ_i^B on observations from those car models k satisfying $k \in \mathcal{A}_{i+1,i}^B \cap \mathcal{A}_j^A$. That is, $k \in \mathcal{A}_{i+1,i}^B \cap \mathcal{A}_{r_k+1}^B \cap \mathcal{A}_{r_k}^A \cap \mathcal{A}_j^A$ with $r_k \in \{j,\ldots,i-1\}$. If one has problems in finding a sufficiently large number of car models satisfying this condition for each i, that is, if (10.72) does not contribute with significantly more information in addition to (10.71) then one can confine oneself to base the estimation on (10.71). By doing so we have to be aware of that we lose some information.

If one really suspects that there is a considerable amount of information in those car models satisfyng (10.72) one should of course utilize this information and base the estimation of ϱ_i^B on car models satisfying(10.72) as well.

When basing the estimation of ϱ_i^B on observations from e.g. only the two last years ($d = 1$) then we automatically base the estimation on only those car models satisfying (10.71).

$\pi_{i+1,j}^B$ $(j = 1,\ldots,i)$ may be computed using (10.66) with A replaced by B.

Time-homogeneous model

From Assumption 6.5 we see that in this model we have

$$\lambda_{i+1,j} = \varrho\lambda_{ij}. \quad (m(i,d) \le j \le i \,;\, i = 1,\ldots,n-1) \tag{10.73}$$

Summing both sides of (10.73) over j and i yields

$$\sum_{i=1}^{n-1} \sum_{j=m(i,d)}^{i} \lambda_{i+1,j} = \sum_{i=1}^{n-1} \sum_{j=m(i,d)}^{i} \varrho\lambda_{ij}$$

implying

$$\varrho = \frac{\sum_{i=1}^{n-1} \sum_{j=m(i,d)}^{i} \lambda_{i+1,j}}{\sum_{i=1}^{n-1} \sum_{j=m(i,d)}^{i} \lambda_{ij}}.$$

Hence, a reasonable estimator of ϱ is given by

$$_d\hat{\varrho} = \frac{\sum_{i=1}^{n-1} \sum_{j=m(i,d)}^{i} \hat{\lambda}_{i+1,j}}{\sum_{i=1}^{n-1} \sum_{j=m(i,d)}^{i} \hat{\lambda}_{ij}} \tag{10.74}$$

where $\hat{\lambda}_{rj}$ $(j = m(i,d), \ldots, i\ ;\ r = i, i+1)$ are given by Algorithm 10.1. We may rewrite (10.74) as

$$_d\hat{\varrho} = \sum_{i=1}^{n-1} u_i^{\varrho}\, {}_d\hat{\varrho}_i$$

with

$$u_i^{\varrho} = \frac{\sum_{j=m(i,d)}^{i} \hat{\lambda}_{ij}}{\sum_{i=1}^{n-1} \sum_{j=m(i,d)}^{i} \hat{\lambda}_{ij}}$$

and ${}_d\hat{\varrho}_i$ is given by (10.65).

10.2.5 An iterative procedure for the estimation of β_j^p and λ_j^p

We have derived estimators of φ_j^p, β_j^p, λ_j^p and ϱ_j^p. Some of the estimators of β_j^p and λ_j^p depend on each other. This is the case for the estimators (10.11) and (10.33). A problem occurs if one wants to use both these estimators, since it is impossible to obtain analytical expressions for both of them. The problem has to be solved numerically. De Vylder(1981) proposes an iterative procedure for handling such estimators. In the present subsection we will pursue this line of thought.

In the lth iteration $(l = 1, 2, \ldots)$ we compute

$$\hat{\beta}_j^{p,(l)} = f\left(\hat{\lambda}_j^{p,\hat{\kappa},(l-1)}\right)$$

given by (10.11), and

$$\hat{\lambda}_j^{p,\hat{\kappa},(l)} = g\left(\hat{\beta}_j^{p,(l)}, \hat{\lambda}_j^{p,\hat{\kappa},(l-1)}\right)$$

given by (10.33). As starting values we use

$$\hat{\beta}_j^{p,(0)} = \dot{\beta}_j^p$$

where $\dot{\beta}_j^p$ is given by (10.12), and

$$\hat{\lambda}_j^{p,\hat{\kappa},(0)} = \hat{\lambda}_j^p$$

where $\hat{\lambda}_j^p$ is given by (10.44) after a restriction of $\check{\lambda}_j^p$ to the interval $[0, \infty)$. The iterative procedure should stop when

$$\left|\hat{\beta}_j^{p,(l)} - \hat{\beta}_j^{p,(l-1)}\right| \quad \text{and} \quad \left|\hat{\lambda}_j^{p,\hat{\kappa},(l)} - \hat{\lambda}_j^{p,\hat{\kappa},(l-1)}\right|$$

become sufficiently small.

10.3 Estimation of future parameters

10.3.1 Introduction

In this section we will propose estimators of the structural parameters β_{n+1}^p, λ_{n+1}^p and ϱ_n^p. The ideas are based on Sundt(1983).

Since we have not yet received any claim data from year $n+1$ we must introduce more structure into the time-heterogeneous models to be able to estimate the above-mentioned parameters. We assume the relevant parameters to be stochastic processes and independent of the Θ_k's. Then all the expectations and covariances introduced in Chapters 5 and 7 will become their conditional analogues given the relevant parameters.

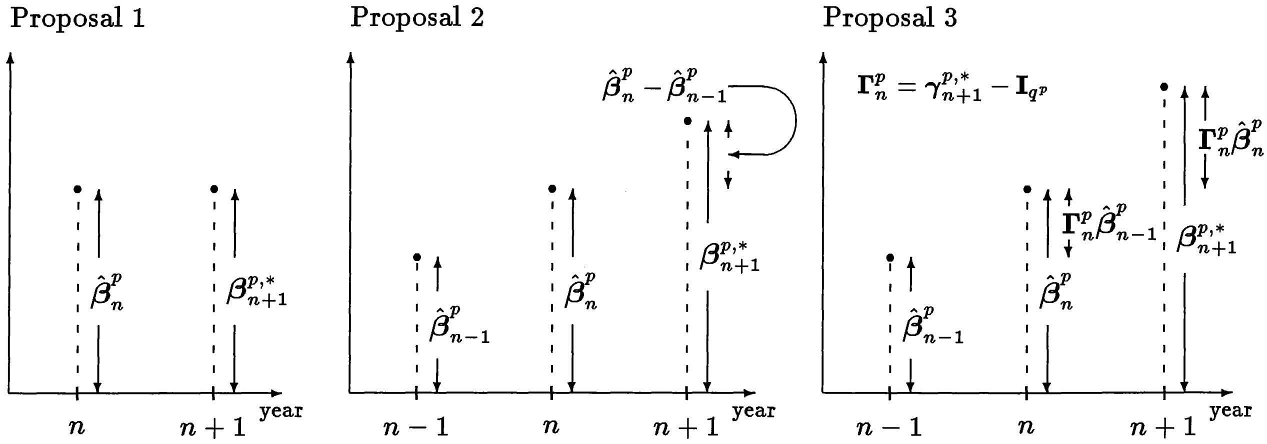

Figure 10.1: Graphical representation of the three proposals of estimators of $\beta^{p,*}_{n+1}$

10.3.2 Martingale-assumptions

Assume that both processes $\{\varrho^p_j\}_{j=1}^{\infty}$ and $\{\lambda^p_j\}_{j=1}^{\infty}$ are martingales. The definition of martingales is (cf. e.g. KARLIN & TAYLOR(1975, p. 238))

$$\mathrm{E}(\varrho^p_n | \varrho^p_1, \ldots, \varrho^p_{n-1}) = \varrho^p_{n-1} \qquad \mathrm{E}(\lambda^p_{n+1} | \lambda^p_1, \ldots, \lambda^p_n) = \lambda^p_n.$$

Hence $\varrho^{p,*}_n = \hat{\varrho}^p_{n-1}$ and $\lambda^{p,*}_{n+1} = \hat{\lambda}^p_n$ are natural estimators of ϱ^p_n and λ^p_{n+1}, respectively.

We assume that we use the same technical variables during the relevant years (of course, their *values* may change over these years, for instance because of increasing price of the car model if price is one of the variables). This implies that $q^p_j = q^p$ $(j = n, n+1)$ in proposal 1 below and $(j = n-1, n, n+1)$ in proposal 2. This assumption does not give any loss of generality. If we have not used the same technical variables during the relevant years (from years n to $n+1$ in proposal 1 and years $(n-1) - (n+1)$ in proposals 2 and 3) then we can just estimate the β^p_j's over again using the same technical variables as in year $n+1$ in all past relevant years.

In order to estimate β^p_{n+1} we have three proposals (see Figure 10.1):

1. Assume that $\{\beta^p_j\}_{j=1}^{\infty}$ are martingales. Then we have
$$\mathrm{E}(\beta^p_{n+1} | \beta^p_1, \ldots, \beta^p_n) = \beta^p_n.$$
$\beta^{p,*}_{n+1} = \hat{\beta}^p_n$ is therefore a reasonable estimator of β^p_{n+1}.

2. Let $\eta^p_j = \beta^p_j - \beta^p_{j-1}$ and assume that $\{\eta^p_j\}_{j=2}^{\infty}$ are martingales. Then $\beta^{p,*}_{n+1} = 2\hat{\beta}^p_n - \hat{\beta}^p_{n-1}$ is a reasonable estimator of β^p_{n+1}.

3. Put $\beta^{p,(q^p \times 1)}_j = (\beta^p_{1j}, \ldots, \beta^p_{qj})'$ $\forall j$. Let
$$\gamma^{p,(q^p \times q^p)}_j = \mathrm{diag}\left\{\beta^p_{1j}/\beta^p_{1,j-1}, \ldots, \beta^p_{q^p j}/\beta^p_{q^p,j-1}\right\} \ \forall j.$$
This gives $\beta^p_j = \gamma^p_j \beta^p_{j-1}$ $\forall j$. Assume that $\{\gamma^p_j\}_{j=2}^{\infty}$ are martingales. Then
$$\mathrm{E}(\gamma^p_{n+1} | \gamma^p_2, \ldots, \gamma^p_n) = \gamma^p_n.$$

This gives that

$$\gamma_{n+1}^{p,*} = \hat{\gamma}_n^p = \operatorname{diag}\left\{\hat{\beta}_{1n}^p/\hat{\beta}_{1,n-1}^p, \ldots, \hat{\beta}_{q^p n}^p/\hat{\beta}_{q^p,n-1}^p\right\}$$

is a reasonable estimator of γ_{n+1}^p . Then

$$\beta_{n+1}^{p,*} = \gamma_{n+1}^{p,*}\dot{\beta}_n^p = \left((\hat{\beta}_{1n}^p)^2/\hat{\beta}_{1,n-1}^p, \ldots, (\hat{\beta}_{q^p n}^p)^2/\hat{\beta}_{q^p,n-1}^p\right)'$$

is a reasonable estimator of β_{n+1}^p.

Notice that proposals 2 and 3 presuppose that we have at hand observations from at least two years back.

To estimate $\mu_{s,n+1}$ we use

$$\mu_{s,n+1}^* = \mathbf{t}_{s,n+1}'\beta_{n+1}^{p,*}.$$

We have been able to find estimators of λ_{n+1},ϱ_n and $\mu_{s,n+1}$ even though we have not yet received any claim data from year $n+1$. To be able to do this we have made some quite strong assumptions. If we really have confidence in these assumptions then we should use these assumptions to estimate the past parameters as well. This has previously been pointed out by SUNDT(1983, p. 189).

As we have seen, there are many ways to estimate the future structural parameters. As a guide for deciding upon which procedure to select in a given situation, one can compute

$$\begin{aligned}
R_n^A\left(\{\varrho_j^{A,*}\}_{j=2}^{n-1}, \{\lambda_j^{A,*}\}_{j=3}^n, \{\beta_j^{A,*}\}_{j=3}^n\right) &= \sum_{j=3}^n \sum_{k\in\mathcal{A}_j^A} w_{kj\cdot}\left(Y_{kj\cdot} - \tilde{m}_{k,j|j-1}^*\right)^2 \\
R_n^B\left(\{\varrho_j^{B,*}\}_{j=2}^{n-1}, \{\varrho_j^{BA,*}\}_{j=1}^{n-2}, \{\lambda_j^{B,*}\}_{j=3}^n, \{\beta_j^{B,*}\}_{j=3}^n\right) &= \sum_{j=3}^n \sum_{k\in\mathcal{A}_j^B} w_{kj\cdot}\left(Y_{kj\cdot} - \tilde{m}_{k,j|j-1}^*\right)^2
\end{aligned}$$

where $\tilde{m}_{k,j|j-1}^*$ $(j = 3, \ldots, n)$ are given by (10.76), (10.79), (10.80), (10.81) and (10.82), and use the procedure which gives the smallest R_n^p-value. (We allow for different procedures in the two sub-portfolios A and B).

10.4 Empirical recursive credibility estimators

10.4.1 Time-heterogeneous models

For the sake of simplicity we introduce $(p = A, B\ ;\ j = 1, \ldots, n)$

$$\begin{aligned}
\hat{V}_{sj} &= \frac{v_{sj\cdot}\hat{\psi}_{s,j|j-1}}{v_{sj\cdot}\hat{\psi}_{s,j|j-1} + \hat{\varphi}_j^p} Y_{sj\cdot} + \frac{\hat{\varphi}_j^p}{v_{sj\cdot}\hat{\psi}_{s,j|j-1} + \hat{\varphi}_j^p}\hat{\tilde{m}}_{s,j|j-1} - \hat{\mu}_{sj} \\
\hat{W}_{sj} &= \frac{\hat{\psi}_{s,j|j-1}\hat{\varphi}_j^p}{v_{sj\cdot}\hat{\psi}_{s,j|j-1} + \hat{\varphi}_j^p} - \hat{\lambda}_j^p.
\end{aligned}$$

Theorem 10.1 *By inserting the estimators of* $\{\beta_j^p\}_{j=1}^{n+1}$, $\{\varphi_j^p\}_{j=1}^n$, $\{\lambda_j^p\}_{j=1}^{n+1}$ *and* $\{\varrho_j^p\}_{j=1}^n$, *developed in Sections 10.2 and 10.3, into the expressions of Theorem 7.1, we get the empirical recursive credibility estimators.*

For $s \in \mathcal{A}^p_{b_s}$ $(p = A, B)$ we have:

$$\hat{\tilde{m}}_{s,b_s|b_s-1} = \hat{\mu}_{sb_s} \qquad \hat{\psi}_{s,b_s|b_s-1} = \hat{\lambda}^p_{b_s} \quad (b_s \le n) \tag{10.75}$$

$$\tilde{m}^*_{s,n+1|n} = \mu^*_{s,n+1} \qquad \psi^*_{s,n+1|n} = \lambda^{p,*}_{n+1}. \quad (b_s = n+1) \tag{10.76}$$

Further, for $s \in \mathcal{A}^A_{b_s}$ we have

$$\begin{aligned}
\hat{\psi}_{s,j+1|j} &= \left(\hat{\varrho}^A_j\right)^2 \hat{W}_{sj} + \hat{\lambda}^A_{j+1} \quad (j = b_s, \dots, \min(r_s, n) - 1) &(10.77)\\
\hat{\psi}_{s,r_s+1|r_s} &= \left(\hat{\varrho}^{BA}_{r_s}\right)^2 \hat{W}_{sr_s} + \hat{\lambda}^B_{r_s+1} \quad (\text{if } r_s \le n-1)\\
\hat{\psi}_{s,j+1|j} &= \left(\hat{\varrho}^B_j\right)^2 \hat{W}_{sj} + \hat{\lambda}^B_{j+1} \quad (\text{if } r_s < n-1\ ;\ j = r_s+1, \dots, n-1)\\
\psi^*_{s,n+1|n} &= \left(\varrho^{B,*}_n\right)^2 \hat{W}_{sn} + \lambda^{B,*}_{n+1} \quad (\text{if } r_s \le n-1)\\
\psi^*_{s,r_s+1|r_s} &= \left(\varrho^{BA,*}_{r_s}\right)^2 \hat{W}_{sr_s} + \lambda^{B,*}_{r_s+1} \quad (\text{if } r_s = n)\\
\hat{\psi}_{s,j+1|j} &= \left(\hat{\varrho}^A_j\right)^2 \hat{W}_{sj} + \hat{\lambda}^A_{j+1} \quad (\text{if } r_s > n\ ;\ j = b_s, \dots, n-1)\\
\psi^*_{s,n+1|n} &= \left(\varrho^{A,*}_n\right)^2 \hat{W}_{sn} + \lambda^{A,*}_{n+1} \quad (\text{if } r_s > n)\\
\hat{\tilde{m}}_{s,j+1|n} &= \hat{\varrho}^A_j \hat{V}_{sj} + \hat{\mu}_{s,j+1} \quad (j = b_s, \dots, \min(r_s, n) - 1) &(10.78)\\
\hat{\tilde{m}}_{s,r_s+1|r_s} &= \hat{\varrho}^{BA}_{r_s} \hat{V}_{sr_s} + \hat{\mu}_{s,r_s+1} \quad (\text{if } r_s \le n-1)\\
\hat{\tilde{m}}_{s,j+1|j} &= \hat{\varrho}^B_j \hat{V}_{sj} + \hat{\mu}_{s,j+1} \quad (\text{if } r_s < n-1\ ;\ j = r_s+1, \dots, n-1)\\
\tilde{m}^*_{s,n+1|n} &= \varrho^{B,*}_n \hat{V}_{sn} + \mu^*_{s,n+1} \quad (\text{if } r_s \le n-1) &(10.79)\\
\tilde{m}^*_{s,r_s+1|r_s} &= \varrho^{BA,*}_{r_s} \hat{V}_{sr_s} + \mu^*_{s,r_s+1} \quad (\text{if } r_s = n) &(10.80)\\
\tilde{m}^*_{s,n+1|n} &= \varrho^{A,*}_n \hat{V}_{sn} + \mu^*_{s,n+1} \quad (\text{if } r_s > n) &(10.81)
\end{aligned}$$

and for $s \in \mathcal{A}^B_{b_s}$ we have

$$\begin{aligned}
\hat{\psi}_{s,j+1|j} &= \left(\hat{\varrho}^B_j\right)^2 \hat{W}_{sj} + \hat{\lambda}^B_{j+1} \quad (j = b_s, \dots, n-1)\\
\psi^*_{s,n+1|n} &= \left(\varrho^{B,*}_n\right)^2 \hat{W}_{sn} + \lambda^{B,*}_{n+1}\\
\hat{\tilde{m}}_{s,j+1|j} &= \hat{\varrho}^B_j \hat{V}_{sj} + \hat{\mu}_{s,j+1} \quad (j = b_s, \dots, n-1)\\
\tilde{m}^*_{s,n+1|n} &= \varrho^{B,*}_n \hat{V}_{sn} + \mu^*_{s,n+1}. &(10.82)
\end{aligned}$$

If the system has been in force before, then we must begin by updating the estimators from the last year we have observations ("$*$-estimators" from year $n+1$ in the previous updating replaced by "$\hat{}$-estimators" from year n in this updating). $\hat{\tilde{m}}_{s,n|n-1}$ and $\hat{\psi}_{s,n|n-1}$ are then computed using these new estimators. Most of the above recursions are in use only when we implement the system for the first time.

10.4.2 Time-homogeneous model

Substitute $\hat{\varphi}_j \equiv \hat{\varphi}$, $\hat{\lambda}_j \equiv \hat{\hat{\lambda}}$, $\hat{\varrho}_j \equiv \hat{\varrho}$ and $\hat{\mu}_{sj} = \hat{\mu}_s$ $\forall j$ into (10.75), (10.77) and (10.78). Then delete A from all the formulae.

Chapter 11

Practical Modifications

11.1 Introduction

The theory developed so far is not completely feasible in practice. The estimator of φ_j^p derived in subsection 10.2.1 is not quite suitable for the data we had at hand in Storebrand. To amend for this problem we introduce a Poisson-assumption on the claim numbers and thus are able to derive an alternative estimator. The credibility estimator derived in Section 10.4 can not be used as a candidate for the rating class; it has the wrong scaling. This is being amended for in Section 11.3. In Section 11.4 we give confidence intervals for the "correct" rating factor and class, under a normal assumption. Finally, in Section 11.5 we give algorithms to simplify the implementation of the system.

For simplicity we are only going to consider the model of Chapter 5 in this chapter. Similar results and expressions can be derived in the models of Chapters 6 and 7 as well.

11.2 Poisson-assumption

As discussed in subsection 3.6 i SUNDT(1987a) there are practical problems associated with our estimator of φ_j. In Storebrand, claim data are not coinciding with policy data . The file of claim data consists of reported claims during the calendar year, whereas the file of policy data originated in the middle of the same year. The registered premium is therefore the premium from the last day it was due prior to the middle of the year. This date may stretch from the middle of the preceding year to the middle of the present year.

Analogously to SUNDT(1987a) we make some additional assumptions to reach at an alternative estimator of φ_j. Let N_{kji} be the number of claims from unit i of car model k in year j and let Z_{kjim} denote the claim amount of the mth of these. Then

$$X_{kji} = \sum_{m=1}^{N_{kji}} Z_{kjim}.$$

We assume that the Z_{kjim}'s are conditionally independent and identically distributed given Θ_{kj} and conditionally independent of the N_{kji}'s. Further, we assume that N_{kji} is conditionally Poisson distributed with parameter $w_{kji}r_{kj}(\Theta_{kj})$ given Θ_{kj}.

From these assumptions we have

$$\mathrm{Var}(X_{kji}|\Theta_{kj}) \quad = \quad \mathrm{E}\left[\mathrm{Var}(X_{kji}|\Theta_{kj}, N_{kji})|\Theta_{kj}\right] + \mathrm{Var}\left[\mathrm{E}(X_{kji}|\Theta_{kj}, N_{kji})|\Theta_{kj}\right]$$

$$
\begin{aligned}
&= \mathrm{E}[\mathrm{Var}(\sum_{m=1}^{N_{kji}} Z_{kjim}|\Theta_{kj}, N_{kji})|\Theta_{kj}] + \mathrm{Var}[\mathrm{E}(\sum_{m=1}^{N_{kji}} Z_{kjim}|\Theta_{kj}, N_{kji})|\Theta_{kj}] \\
&\stackrel{(1)}{=} \mathrm{E}[\sum_{m=1}^{N_{kji}} \mathrm{Var}(Z_{kjim}|\Theta_{kj}, N_{kji})|\Theta_{kj}] + \mathrm{Var}[\sum_{m=1}^{N_{kji}} \mathrm{E}(Z_{kjim}|\Theta_{kj}, N_{kji})|\Theta_{kj}] \\
&\stackrel{(2)}{=} \mathrm{E}(N_{kji}|\Theta_{kj})\mathrm{Var}(Z_{kjim}|\Theta_{kj}) + [\mathrm{E}(Z_{kjim}|\Theta_{kj})]^2\,\mathrm{Var}(N_{kji}|\Theta_{kj}) \\
&\stackrel{(3)}{=} w_{kji}r_{kj}(\Theta_{kj})\left\{\mathrm{Var}(Z_{kjim}|\Theta_{kj}) + [\mathrm{E}(Z_{kjim}|\Theta_{kj})]^2\right\} \\
&= w_{kji}r_{kj}(\Theta_{kj})q_{kj}(\Theta_{kj})
\end{aligned}
$$

where

$$q_{kj}(\Theta_{kj}) = \mathrm{E}(Z^2_{kjim}|\Theta_{kj}).$$

(1) follows from the fact that the Z_{kjim}'s are conditionally independent given Θ_{kj}. (2) is a consequence of the Z_{kjim}'s being conditionally independent of the N_{kji}'s for given Θ_{kj} and identically distributed. (3) is due to the fact that the N_{kji}'s are conditionally Poisson distributed given Θ_{kj}.

On the other hand we have from the definition of Y_{kji} that

$$
\begin{aligned}
\mathrm{Var}(X_{kji}|\Theta_{kj}) &= \mathrm{Var}(w_{kji}Y_{kji}|\Theta_{kj}) = w^2_{kji}\mathrm{Var}(Y_{kji}|\Theta_{kj}) = w^2_{kji}\frac{s^2_j(\Theta_{kj})}{v_{kji}} \\
&= w_{kji}s^2_j(\Theta_{kj}),
\end{aligned}
$$

where the last equality follows from $v_{kji} = w_{kji}$. This implies that

$$s^2_j(\Theta_{kj}) = r_{kj}(\Theta_{kj})q_{kj}(\Theta_{kj}). \tag{11.1}$$

Then

$$\check{\varphi}_{kj} = \frac{\sum_{i=1}^{I_{kj}}\sum_{m=1}^{N_{kji}} Z^2_{kjim}}{w_{kj\cdot}}$$

is a Θ_{kj}-unbiased estimator of $s^2_j(\Theta_{kj})$. To see this we consider

$$
\begin{aligned}
\mathrm{E}(\check{\varphi}_{kj}|\Theta_{kj}) &= \mathrm{E}\{\mathrm{E}[\frac{\sum_{i=1}^{I_{kj}}\sum_{m=1}^{N_{kji}} Z^2_{kjim}}{w_{kj\cdot}}|\Theta_{kj}, N_{kji}]|\Theta_{kj}\} \\
&= \frac{1}{w_{kj\cdot}}\mathrm{E}[\sum_{i=1}^{I_{kj}}\sum_{m=1}^{N_{kji}} \mathrm{E}(Z^2_{kjim}|\Theta_{kj})|\Theta_{kj}] = \frac{1}{w_{kj\cdot}}\left[\sum_{i=1}^{I_{kj}} \mathrm{E}(N_{kji}|\Theta_{kj})\right] q_{kj}(\Theta_{kj}) \\
&= \frac{1}{w_{kj\cdot}}\left[\sum_{i=1}^{I_{kj}} w_{kji}r_{kj}(\Theta_{kj})\right] q_{kj}(\Theta_{kj}) = \frac{1}{w_{kj\cdot}}w_{kj\cdot}s^2_j(\Theta_{kj}) = s^2_j(\Theta_{kj}).
\end{aligned}
$$

Thus

$$\check{\varphi}_j = \sum_{k\in\mathcal{A}_j} u^{\varphi}_{kj}\check{\varphi}_{kj} \tag{11.2}$$

is an unbiased estimator of φ_j for all weights u^{φ}_{kj} such that $\sum_{k\in\mathcal{A}_j} u^{\varphi}_{kj} = 1$.

From (11.1) we see that r_{kj} and q_{kj} depend on k, whereas the product is independent of k. On the other hand the technical data should influence on the number of claims. We must therefore make some modifications. Numerical studies in SUNDT(1987a) indicate that $\check{\varphi}_{kj}$ is correlated with the technical variables weight, engine power, cylinder volume and price. But we have that $\check{\varphi}_{kj}/e_{kj}$, where e_{kj}=engine power for car model k in year j, is significantly less correlated with the same technical variables. We will therefore replace assumption $\mathrm{Var}(Y_{kji}|\Theta_{kj}) = s_j^2(\Theta_{kj})/w_{kji}$ by $\mathrm{Var}(Y_{kji}|\Theta_{kj}) = e_{kj}s_j^2(\Theta_{kj})/w_{kji} = s_j^2(\Theta_{kj})/v_{kji}$, where $v_{kji} = w_{kji}/e_{kj}$. Under this assumption we replace (11.1) by

$$s_j^2(\Theta_{kj}) = \frac{r_{kj}(\Theta_{kj})q_{kj}(\Theta_{kj})}{e_{kj}}.$$

We then get that

$$\check{\varphi}_{kj} = \frac{\sum_{i=1}^{I_{kj}} \sum_{m=1}^{N_{kji}} Z_{kjim}^2}{e_{kj}w_{kj\cdot}}$$

is a Θ_{kj}-unbiased estimator of $s_j^2(\Theta_{kj})$. Thus

$$\check{\varphi}_j = \sum_{k\in\mathcal{A}_j} u_{kj}^{\varphi}\check{\varphi}_{kj}$$

is an unbiased estimator of φ_j. Finally we replace $\hat{\varphi}_j$ by $\check{\varphi}_j$ in all the expressions in the time-heterogeneous model.

All of the derivations in the time-heterogeneous model are still valid under the modified assumptions introduced in this section. But we have to remember that $s_j^2(\Theta_{kj})$, φ_j and v_{kji} have different interpretations, while $s_j^2(\Theta_{kj})/v_{kji}$ and φ_j/v_{kji} (and hence $s_j^2(\Theta_{kj})/v_{kj\cdot}$ and $\varphi_j/v_{kj\cdot}$ as well) are unchanged.

11.3 Estimated rating factor

In Section 10.4 we have derived expressions for the empirical recursive credibility estimators $\left\{\tilde{m}_{s,t+1|t}^*\right\}_{t=b_s-1}^{\infty}$ and their respective estimated estimation errors $\left\{\psi_{s,t+1|t}^*\right\}_{t=b_s-1}^{\infty}$. But we can still not use $\tilde{m}_{s,n+1|n}^*$ as a candidate for the rating factor of car model s in year $n+1$. It has to be adjusted by a scaling factor. We will use the same arguments as in subsection 3.4 of SUNDT(1987a). Here the scaling factor is calculated under the assumption that the total premium in year n should be the same with the new rating factor as with the old one. If we let γ_{n+1}^* be our scaling factor then our new factor of car model s in year $n+1$ will be $f_{s,n+1}^* = \gamma_{n+1}^*\tilde{m}_{s,n+1|n}^*$. The above-mentioned criterion gives that the scaling factor is given by

$$\gamma_{n+1}^* = \frac{\sum_{k\in\mathcal{A}_n} p_{kn\cdot}}{\sum_{k\in\mathcal{A}_n} w_{kn\cdot}\tilde{m}_{k,n+1|n}^*}$$

where $p_{kn\cdot} = \sum_{i=1}^{I_{kn}} p_{kni}$.

This leads to our proposition $k_{s,n+1}^*$ for the rating class of car model s in year $n+1$ for vehicle damage insurance:

- The integer between 30 and 94 closest to $30 + \log(f_{s,n+1}^*)/\log(1.04)$.

From Chapter 1 we remember that the factors for the 6 classes in liability insurance are 0.75, 1.00, 1.07, 1.13, 1.33 and 1.50, respectively. It is natural to choose these factors as the approximate midpoints for the 6 classes when attaining a class to each factor proposed. Hence our proposition $k_{s,n+1}^*$ for the rating class of car model s in year $n+1$ for liability insurance equals:

- 1 if $f^*_{s,n+1} \in (-\infty, 0.9]$, 2 if $f^*_{s,n+1} \in (0.9, 1.04]$, 3 if $f^*_{s,n+1} \in (1.04, 1.1]$, 4 if $f^*_{s,n+1} \in (1.1, 1.25]$, 5 if $f^*_{s,n+1} \in (1.25, 1.4]$, and 6 if $f^*_{s,n+1} \in (1.4, \infty)$.

The final classification $k_{s,n+1}$ is made by a person with a thorough knowledge about the different car models. This person compares $k^*_{s,n+1}$ with the classification of the same car model from previous years (if it is not new), the classification of similar car models this year, the classification of the same car model in competing companies and so on. Thus, the final rating factor of car model s in year $n+1$ is given by

$$f_{s,n+1} = 1.04^{k_{s,n+1}-30}$$

for vehicle damage and the obvious analogue for liability insurance.

11.4 Confidence intervals for the "correct" rating factor and class

In this section we are going to derive a confidence interval for the "correct" rating factor $\gamma_{n+1} m_{s,n+1}(\Theta_{s,n+1})$, where $\gamma_{n+1} = \mathrm{E}(\gamma^*_{n+1})$. This confidence interval is supposed to be a guide for the person who is going to make the final classification of the car model. In this way he is able to decide to what extent he can rely on the proposed estimate.

As in the previous section we will follow the same lines as the ones in subsection 3.4 of SUNDT(1987a). We assume that the conditional distribution of $m_{s,n+1}(\Theta_{s,n+1})$ given the observations is approximately normal with expectation $\tilde{m}^*_{s,n+1|n}$ and variance $\psi^*_{s,n+1|n}$. This gives

$$\Pr\left(-g_{1-\varepsilon/2} \leq \frac{m_{s,n+1}(\Theta_{s,n+1}) - \tilde{m}^*_{s,n+1|n}}{\sqrt{\psi^*_{s,n+1|n}}} \leq g_{1-\varepsilon/2} | \mathbf{Y}\right) \approx 1 - \varepsilon,$$

where $g_{1-\varepsilon/2}$ is the $1-\varepsilon/2$ fractile of the standard normal distribution. In other words

$$\Pr\left(\tilde{m}^*_{s,n+1|n} - g_{1-\varepsilon/2}\sqrt{\psi^*_{s,n+1|n}} \leq m_{s,n+1}(\Theta_{s,n+1}) \leq \tilde{m}^*_{s,n+1|n} + g_{1-\varepsilon/2}\sqrt{\psi^*_{s,n+1|n}} | \mathbf{Y}\right) \approx 1 - \varepsilon. \tag{11.3}$$

Multiplying $\gamma_{n+1} (\geq 0)$ through the inequalities of the left hand side of (11.3) gives

$$\Pr\left(\gamma_{n+1}\tilde{m}^*_{s,n+1|n} - \gamma_{n+1} g_{1-\varepsilon/2}\sqrt{\psi^*_{s,n+1|n}} \leq \gamma_{n+1} m_{s,n+1}(\Theta_{s,n+1}) \right.$$
$$\left. \leq \gamma_{n+1}\tilde{m}^*_{s,n+1|n} + \gamma_{n+1} g_{1-\varepsilon/2}\sqrt{\psi^*_{s,n+1|n}} | \mathbf{Y}\right) \approx 1 - \varepsilon.$$

In the end points of the interval we replace γ_{n+1} by its unbiased estimator γ^*_{n+1}. This gives the following $(1-\varepsilon)100\%$ estimated confidence interval for the "correct" rating factor $\gamma_{n+1} m_{s,n+1}(\Theta_{s,n+1})$, as an approximation:

$$f^*_{s,n+1} \pm \gamma^*_{n+1} g_{1-\varepsilon/2}\sqrt{\psi^*_{s,n+1|n}}.$$

An at least $(1-\varepsilon)100\%$ estimated confidence interval for the rating class of car model s in year $n+1$ is approximated by (vehicle damage):

- Lower bound is given by the greatest integer between 30 and 94 less than or equal to $30 + \log(f^*_{s,n+1} - \gamma^*_{n+1} g_{1-\varepsilon/2}\sqrt{\psi^*_{s,n+1|n}})/\log(1.04)$.

- Upper bound is given by the lowest integer between 30 and 94 greater than or equal to $30 + \log(f^*_{s,n+1} + \gamma^*_{n+1} g_{1-\varepsilon/2}\sqrt{\psi^*_{s,n+1|n}})/\log(1.04)$

and the obvious analogues for liability insurance.

11.5 Preparations for the implementation of the system

11.5.1 Introduction

In this section we are going to determine which data from previous years we have to store from year to year to be able to estimate the structural parameters. We will give algorithms for the calculation of various statistics which leads to our proposed rating classes for the relevant car models.

11.5.2 Implementation for the first time

We denote the first year we observe the portfolio by year 1. We are going to update the rating structure for year $n+1$. For this purpose we have received observations from n years (this quantity will increase by 1 from one updating to the next).

For the updating procedure we need the following observables $(j = 1, \dots, n)$:

(i) Y_{kji}, w_{kji}, N_{kji}, Z_{kjim}, q_j $\quad (m = 1, \dots, N_{kji}\ ;\ i = 1, \dots, I_{kj}\ ;\ \forall k \in \mathcal{A}_j)$

(ii) I_{kj}, $\mathbf{t}_{kj}$ $\quad (\forall k \in \mathcal{A}_j)$

(iii) $\mathcal{A}_j$

In addition we need to know

(iv) which car models are included in the rating structure in year $n+1$, $\mathbf{t}_{k,n+1}$ from all the car models in the rating structure in year $n+1$, and $p_{kni\cdot}$ $(i = 1, \dots, I_{kn}\ ;\ \forall k \in \mathcal{A}_n)$

We use these basic observables to calculate the statistics (the notation 11.2.6 stands for step 6 of Algorithm 11.2 and so on)

Algorithm 11.1 *For $(j = 1, \dots, n)$ calculate:*

1. $v_{kj\cdot}$, $Y_{kj\cdot}$, $w_{kj\cdot}$, e_{kj} $(\forall k \in \mathcal{A}_j)$ *and* $v_{\cdot j\cdot}$ *by (i), (ii) and (iii)*
2. $\check{\varphi}_j$ *by (i), (ii), (iii), 11.1.1 and* u^{φ}_{kj} $(\forall k \in \mathcal{A}_j)$
3. $\mathbf{D}_j$ *by 11.1.1 and* $\mathbf{T}_j$ *by (ii)*
4. $\mathbf{G}_j = (\mathbf{T}_j'\mathbf{D}_j\mathbf{T}_j)^{-1}\mathbf{T}_j'\mathbf{D}_j$ *by 11.1.3*
5. $\mathbf{Y}_{\Sigma j}$ *by (iii) and 11.1.1*
6. $\dot{\boldsymbol{\beta}}_j$ *by 11.1.4 and 11.1.5*
7. Q_j *by 11.1.3, 11.1.5, 11.1.6 and* $\mathbf{U}^{\lambda}_j$
8. K_j *by (iii)*
9. $\hat{\lambda}_j$ *by (i), (iii), 11.1.1, 11.1.2, 11.1.3, 11.1.4, 11.1.7 and 11.1.8*
10. $\hat{\kappa}_j$ *by 11.1.2 and 11.1.9*
11. $\mathbf{D}^{\hat{\kappa}}_j$ *by (iii), 11.1.1 and 11.1.10*

12. $\hat{\beta}_j$ *by 11.1.3, 11.1.5 and 11.1.11*

Now, choose the number of years we want to base the estimation of ϱ_j on. This number is defined to be $d+1$. Then

Algorithm 11.2 *For* $(i = 1, \ldots, n-1\ ;\ j = m(i,d), \ldots, i\ ;\ r = j, i+1)$ *calculate:*

1. ${}_j^{i+1}\mathbf{Y}_{\Sigma r}$ *by (iii) and 11.1.1*
2. ${}_j^{i+1}\mathbf{T}_r$ *by (ii) and (iii)*
3. $Q_{i+1,j}$ *by 11.1.6, 11.2.1, 11.2.2 and* $\mathbf{U}^{\varrho}_{i+1,j}$
4. $\mathbf{\Delta}^{(l)}_{i+1,j}$ $(l = 1,2,3)$ *by (iii)*
5. ${}_j^{i+1}\mathbf{A}_r$ *by 11.1.4 and 11.2.2*
6. $a_{i+1,j}$ *by 11.2.4, 11.2.5 and* $\mathbf{U}^{\varrho}_{i+1,j}$
7. $\hat{\lambda}_{i+1,j}$ *by Algorithm 10.1*
8. ${}_d\hat{\varrho}_i$ *by 11.2.7*

We are now able to calculate our proposal for the estimators of the rating factor and class.

Algorithm 11.3 *Calculate for all car models k in the rating structure in year* $n+1$

1. $\lambda^*_{n+1}, \beta^*_{n+1}$ *and* ϱ^*_n *by 11.1.9, 11.1.12 and 11.2.8, respectively*
2. $\mu^*_{k,n+1}$ *by (iv) and 11.3.1*
3. $\psi^*_{k,n+1|n}$ *and* $\tilde{m}^*_{k,n+1|n}$ *by (ii), 11.1.1, 11.1.2, 11.1.9, 11.1.12, 11.2.8, 11.3.1 and 11.3.2*
4. γ^*_{n+1} *by (ii), (iii), (iv), 11.1.1 and 11.3.3*
5. $f^*_{k,n+1}$ *by 11.3.3 and 11.3.4*
6. $k^*_{k,n+1}$ *by 11.3.5*
7. *estimated* $(1-\varepsilon)100\%$ *confidence interval by 11.3.3, 11.3.4, 11.3.5 and* $g_{1-\varepsilon/2}$

11.5.3 Later updating

We assume that the system has been used at least once before and is thus initiated. In addition we assume that we have stored the following observables from the last updating:

11.5.3.1. $\mathbf{G}_{n-1}$

11.5.3.2. $\mathbf{t}_{k,n-1}$ $(\forall k \in \mathcal{A}_n)$

11.5.3.3. $\mathcal{A}_{n-1}$

11.5.3.4. The weights from year $n-1$ belonging to $\mathbf{U}^{\varrho}_{n,n-1}$

11.5.3.5. $\mathbf{Y}_{\Sigma,n-1}$

11.5.3.6. $\hat{\lambda}_{n-1,j} \quad (j = m(n-1,d), \ldots, n-1)$

11.5.3.7. $\hat{V}_{k,n-1} = \frac{v_{k,n-1,\cdot}\hat{\psi}_{k,n-1|n-2}}{v_{k,n-1,\cdot}\hat{\psi}_{k,n-1|n-2}+\hat{\varphi}_{n-1}} Y_{k,n-1,\cdot} + \frac{\hat{\varphi}_{n-1}}{v_{k,n-1,\cdot}\hat{\psi}_{k,n-1|n-2}+\hat{\varphi}_{n-1}} \hat{\hat{m}}_{k,n-1|n-2} - \hat{\mu}_{k,n-1}$
$(\forall k \in \mathcal{A}_n)$

11.5.3.8. $\hat{W}_{k,n-1} = \frac{\hat{\psi}_{k,n-1|n-2}\hat{\varphi}_{n-1}}{v_{k,n-1,\cdot}\hat{\psi}_{k,n-1|n-2}+\hat{\varphi}_{n-1}} - \hat{\lambda}_{n-1} \quad (\forall k \in \mathcal{A}_n)$

We need these observables from year n:

(I) $Y_{kni}, w_{kni}, N_{kni}, Z_{knim} \quad (m = 1, \ldots, N_{kni}\ ;\ i = 1, \ldots, I_{kn}\ ;\ \forall k \in \mathcal{A}_n)$

(II) $I_{kn}, \mathbf{t}_{kn} \quad (\forall k \in \mathcal{A}_n)$

(III) $\mathcal{A}_n$

In addition we need to know

(IV) which car models are included in the rating structure in year $n+1$, $\mathbf{t}_{k,n+1}$ from all the car models in the rating structure in year $n+1$, and p_{kni}. $(i = 1, \ldots, I_{kn}\ ;\ \forall k \in \mathcal{A}_n)$

We now use Algorithm 11.1 with $j = n$ and replace (i)–(iv) by (I)–(IV). Then we use Algorithm 11.2 with $j = n-1$.

Algorithm 11.4 *For all $k \in \mathcal{A}_n$ calculate*

1. *$\hat{\mu}_{kn}$ by (II) and 11.1.12*
2. *$\hat{\hat{m}}_{k,n|n-1}$ by 11.5.3.7, 11.2.8 and 11.4.1*
3. *$\hat{\psi}_{k,n|n-1}$ by 11.5.3.8, 11.1.9 and 11.2.8*

Finally we use Algorithm 11.3 with (i)–(iv) replaced by (I)–(IV).

Chapter 12

Application on Real Data

12.1 Introduction

The time-heterogeneous model (Chapter 5) was implemented during the summer of 1990 using data from Storebrand for the three year period 1987–1989.

We assume that we are at the end of 1989 and want to make a classification of the car models for 1990. To simplify notation, 1987 will be denoted by year 1, 1988 year 2, and so on. To compute the credibility estimator of a car model we have to know the technical variables of that car model. Gathering information about the engine powers and the weights of all the car models in the portfolio would have been too expensive and time-consuming for this project. Besides, the prices are not known for all the car models in the portfolio since many of them are no longer for sale. Thus we decided to take a sample of the car models in the portfolio and compute the credibility estimator for these car models. The engine power, weight and price of each of these car models was for 1987 taken from a list from November 1987 (Opplysningsrådet for Veitrafikken(1987)), for 1988 a list from February 1988 (Opplysningsrådet for Veitrafikken(1988)), and for 1989 a list from May 1990 (Opplysningsrådet for Veitrafikken(1990)). We only included car models that were in these lists.

We used the program package SAS (Statistical Analysis System) which was well-suited for arranging all the data we needed for our analysis. Most of the calculations were performed using a module in SAS called IML (Interactive Matrix Language). This module is very easy to use and the program statements are written by almost copying the matrix algebra notation. But problems did occur during the parameter estimation process. The data system was not able to allocate sufficient memory to the main storage to be able to perform the most ponderous matrix manipulations. This problem was solved by storing the largest matrices on disk and load into the main storage only the matrices needed for each calculation. In the beginning we used version 5.18 of the SAS system. At the worst we did not have enough space in the main storage to perform a multiplication of two matrices with dimensions 388×370 and 370×344 using this version. The problem was not the multiplication procedure itself, but we were not able to keep both matrices in the main storage at the same time. We therefore had to do the multiplication by loading a row from the first matrix and a coloumn from the second matrix, one at a time. This made the programming bothersome, but it worked. I am not sure whether this is caused by SAS or not. This could very well be a consequence of the operating system which allocates a certain amount of memory to each user of the data system. At a later stage we changed from this version to

year	$\check{\varphi}_j$ (11.2)	$\check{\lambda}_j$ (10.47)	$\check{\lambda}_j^{\hat{\kappa}}$ (10.31)	$\dot{\boldsymbol{\beta}}_j$ (10.12)	$\hat{\boldsymbol{\beta}}_j$ (10.11)
				-0.403245	-0.483031
1	177 733.95	0.2753803	0.4940231	0.0142767	0.0155997
				0.004672	0.0046979
				-0.594316	-0.67992
2	163 727.73	0.3917311	0.8368299	0.0123948	0.0135615
				0.0063159	0.0064326
				-0.387578	-0.488651
3	209 187.45	0.2167995	0.4292325	0.0139309	0.0160878
				0.0021017	0.0017355

Table 12.1: The original estimators of φ_j, λ_j and $\boldsymbol{\beta}_j$ (vehicle damage).

version 6.06 and all the computations presented in this chapter are made with the latter version. Then it was sufficient for us to make sure that the data system did not have more than three matrices in the main storage at the same time.

12.2 Estimation of the structural parameters

12.2.1 Estimators based on a sample of the car models

We did not base the estimation on all the car models in the portfolio. As with the credibility estimator, for the estimation of $\boldsymbol{\beta}_j$, λ_j and ϱ_j we need technical variables like price, weight and engine power from all the car models to be included in the estimation process. Therefore we decided to take a sample of car models from the portfolio in each year for the estimation of the parameters as well. In addition, we did not include car models introduced to the market in the second half of the year. From these car models we may have claims, but the risk volumes will be zero. From the estimation process we also excluded the most expensive cars (in excess of NOK 750 000) and car models with one or no policies associated with them. Hence, the estimation of φ_j, $\boldsymbol{\beta}_j$ and λ_j are based on observations from 368, 387 and 298 car models in year 1, 2 and 3, respectively. The estimation of ϱ_j is based on observations from the following number of car models: ${}^2_1K = 341$ and ${}^3_2K = 228$. In addition we have ${}^3_1K = 201$.

We estimate φ_j by (11.2) with weights $u^{\varphi}_{kj} = 1/K_j$. The correlation between $\check{\varphi}_{kj}$ and the variables weight, engine power, and price were not considerably higher than the correlation between $\check{\varphi}_{kj}/e_{kj}$ and the same variables. The reductions were only approximately 50%. SUNDT(1987a) observed a much larger reduction. We therefore decided to put $v_{kji} \equiv w_{kji}$. $\boldsymbol{\beta}_j$ is estimated by (10.11) and (10.12), λ_j by (10.31) with weights $\mathbf{U}^{\lambda}_j = \mathbf{D}^{\hat{\kappa}}_j$ and (10.47). We estimate ϱ_j with $d = 1$ (observations from years $j+1$ and j only).

Vehicle damage insurance

The numerical values of these estimators are presented in Table 12.1. The regressors of the estimators of $\boldsymbol{\beta}_j$ are in the tables from the top and down: a constant, engine power and price/weight. The regressor price/weight is the one of the three that is the least

year	1	2
${}_1\hat{\varrho}_j$	1.1926901	0.624943

${}_1\hat{\pi}_{21}$	${}_1\hat{\pi}_{32}$
1	0.84

Table 12.2: The original estimators of ϱ_j, π_{21} and π_{32} (vehicle damage).

stable. From year 1 to year 2 the increase is of 35% in $\dot{\boldsymbol{\beta}}_j$. From year 2 to year 3 there is a reduction of nearly 67%! In $\hat{\boldsymbol{\beta}}_j$ the fluctuation of the regressor price/weight is even more extreme. When the regressor price/weight increases the other two regressors decrease and vice versa. For ordinary car models the increase in the values of both the regressors constant and engine power is not enough to compensate for the reduction of the regressor price/weight from year 2 to year 3. This will give a lower prior estimate in year 3 than in year 2. Notice how stable the values of the regressor engine power are, compared to the other two regressors. It is only natural that the coefficient of engine power is more stable than that of price/weight, since the price is influenced by market conditions, inflation and so on. The engine power of a car model can vary because of technical adjustments of the engine. So the fluctuation of price/weight will have a totally different character than that of engine power. That the prices have been observed in November for 1987, in February for 1988 and in May 1990 for 1989 is clearly a disadvantage. The time-period between the prices in 1987 and 1988 is too short, and between 1988 and 1989 too long. This is probably one of the reasons that the regressor price/weight varies that much from 1988 to 1989; the prices used in 1989 are higher than they should be. We should of course use prices taken from the same month in each year.

The large fluctuation of the estimates of the $\boldsymbol{\beta}_j$'s could be amended for by incorporating an inflation assumption on the $\boldsymbol{\beta}_j$'s. The inflation could be estimated from some sort of exogeneous price index of cars. This will not be done in this paper.

There is a striking difference between the values of $\check{\lambda}_j$ and $\check{\lambda}_j^{\kappa}$. The reason for this difference is given later in this section in connection with the comments to Table 12.3. The values of the estimator of the ϱ_j's are presented in Table 12.2. In the same table are the values of the estimator (10.64) of π_{21} and π_{32} given.

As we see, the other estimators also vary quite a lot over the three year period. For instance $\check{\lambda}_j$ has a reduction of nearly 45% from year 2 to year 3. This is not good. To the present author it does not seem reasonable that the portfolio should change to the extent that is reflected in the parameter estimates. We will therefore modify the estimators to be able to restrict their variation. This will be done with $\check{\varphi}_j$ first.

From Assumptions 5.3 and 5.4 we see that φ_j will be influenced by inflation since the v_{kji}'s are. This is also reflected in the estimator (11.2). To reduce the fluctuation of the φ_j's we assume that they are influenced by inflation only and that the inflation is constant over the three year period. We are willing to make these rather unrealistic assumptions since the observational period is relatively short.

In mathematical terms we can state our assumptions as

$$\varphi_{j+1} = \alpha\varphi_j. \quad (j = 1, 2) \tag{12.1}$$

We have to estimate the inflation or, equivalently, α from our data. Summing both sides of (12.1) over j yields

$$\sum_{j=1}^{2} \varphi_{j+1} = \alpha \sum_{j=1}^{2} \varphi_j$$

or equivalently

$$\alpha = \frac{\sum_{j=1}^{2} \varphi_{j+1}}{\sum_{j=1}^{2} \varphi_j}.$$

Thus, a reasonable estimator of α is given by

$$\hat{\alpha} = \frac{\sum_{j=1}^{2} \check{\varphi}_{j+1}}{\sum_{j=1}^{2} \check{\varphi}_j}, \tag{12.2}$$

with $\check{\varphi}_j$ $(j = 1, 2, 3)$ given by the first column of Table 12.1. The inflation rate i is estimated by

$$\hat{\imath} = \hat{\alpha} - 1.$$

Since the φ_j's vary from year to year because of inflation only, the estimator of each φ_j should depend on observations from all of the three years. This will be done by invoking the inflation assumption in the estimation procedure, which should lead to new estimators.

From (12.1) we have that

$$\breve{\varphi}_{j+1} = \hat{\alpha} \breve{\varphi}_j \quad (j = 1, 2) \tag{12.3}$$

is a reasonable relationship between our new estimators. (12.1) justifies the following scaling of the new estimators

$$\sum_{j=1}^{3} \check{\varphi}_j = \sum_{j=1}^{3} \hat{\alpha}^{j-1} \breve{\varphi}_1.$$

Solving with respect to $\breve{\varphi}_1$ we obtain

$$\breve{\varphi}_1 = \frac{\sum_{j=1}^{3} \check{\varphi}_j}{\sum_{j=1}^{3} \hat{\alpha}^{j-1}}$$

and using (12.3) we have the estimators

$$\breve{\varphi}_i = \hat{\alpha}^{i-1} \frac{\sum_{j=1}^{3} \check{\varphi}_j}{\sum_{j=1}^{3} \hat{\alpha}^{j-1}} = \frac{\sum_{j=1}^{3} \check{\varphi}_j}{\sum_{j=1}^{3} \hat{\alpha}^{j-i}}, \quad (i = 1, 2, 3) \tag{12.4}$$

where $\check{\varphi}_j$ $(j = 1, 2, 3)$ are given by Table 12.1 and $\hat{\alpha}$ by (12.2). We may rewrite (12.2) as

$$\hat{\alpha} = \sum_{j=1}^{2} \frac{\check{\varphi}_j}{\sum_{j'=1}^{2} \check{\varphi}_{j'}} \frac{\check{\varphi}_{j+1}}{\check{\varphi}_j}$$

which is a weighted mean of $\check{\varphi}_2/\check{\varphi}_1$ and $\check{\varphi}_3/\check{\varphi}_2$ with weights $\check{\varphi}_1/(\sum_{j=1}^{2} \check{\varphi}_j)$ and $\check{\varphi}_2/(\sum_{j=1}^{2} \check{\varphi}_j)$, respectively. From Table 12.1 and estimator (12.2) we get $\hat{\alpha} = 1.092$, that is, an estimated inflation rate during the observational period of 9.2%.

We will now have new values of the estimators of the other structural parameters, since they all depend on the estimator of φ_j. These new estimates together with $\breve{\varphi}_j$ defined by (12.4) are given in Table 12.3 and 12.4.

The estimates of λ_j are now fairly stable, but $\check{\lambda}_j$ still has a considerable drop in year 3. The large difference between the estimates of $\check{\lambda}_j$ and $\lambda_j^{\hat{\kappa}}$ is worth investigating more closely.

year	$\breve{\varphi}_j$ (12.4)	$\breve{\lambda}_j$ (10.47)	$\breve{\lambda}_j^{\hat{\kappa}}$ (10.31)	$\dot{\boldsymbol{\beta}}_j$ (10.12)	$\hat{\boldsymbol{\beta}}_j$ (10.11)
1	167 634.09	0.3132175	0.5887911	-0.403245 0.0142767 0.004672	-0.493658 0.0157644 0.0047051
2	183 075.58	0.329973	0.6560925	-0.594316 0.0123948 0.0063159	-0.667192 0.0132647 0.0064708
3	199 939.46	0.249689	0.5180039	-0.387578 0.0139309 0.0021017	-0.503887 0.0163692 0.0016989

Table 12.3: Estimators of φ_j, λ_j and $\boldsymbol{\beta}_j$ where the estimators of φ_j are adjusted for inflation (vehicle damage).

year	1	2
${}_1\hat{\varrho}_j$	1.0263989	0.741908

${}_1\hat{\pi}_{21}$	${}_1\hat{\pi}_{32}$
1	0.85

Table 12.4: Estimators of ϱ_j, π_{21} and π_{32} where the estimators of φ_j are adjusted for inflation (vehicle damage).

If we substitute $\mathbf{D}_j^{\hat{\kappa}}$ by $\mathbf{D}_j$ in the right hand side of (10.38) we get the right hand side of (10.47). Hence $\breve{\lambda}_j$ and $\breve{\lambda}_j^{\hat{\kappa}}$ are equal if $\mathbf{D}_j^{\hat{\kappa}} = \mathbf{D}_j$.

In practice one has to decide which of the two estimators $\breve{\lambda}_j$ and $\breve{\lambda}_j^{\hat{\kappa}}$ to use. Let us investigate the impact these estimators have on the empirical credibility estimator. For convenience we will look at the credibility estimator (5.6). We see that a higher value of the estimator of λ_j will give more weight to the observation from the last year and less weight to the old credibility estimator. In this case the weight given to the past experience could be lower than is wanted. This effect is enforced by the estimator of ϱ_j since from (10.63) we have that ${}_1\dot{\varrho}_j$ is decreasing in the estimator of λ_j ($\hat{\lambda}_{j+1,j}$ is independent of estimators of the structural parameters). So if one wants to estimate $m_{s,t+1}(\Theta_{s,t+1})$ and the estimators of the λ_j's are higher, we would give relatively little weight to the information lying in the years $b_s, \ldots, t$ (possibly with an exception of $Y_{st\cdot}$) and the prior estimator in year $t+1$ will get a larger influence on the estimator of $m_{s,t+1}(\Theta_{s,t+1})$ (this can be unfortunate since a higher value of the estimators of the λ_j's will give a higher estimate of the estimation error of the prior estimator). If this is wanted in practice then one should use $\breve{\lambda}_j^{\hat{\kappa}}$ as the estimator of λ_j. But if one believes that the λ_j's are over-estimated by the $\breve{\lambda}_j^{\hat{\kappa}}$'s, then one should use $\breve{\lambda}_j$. We know that $\breve{\lambda}_j$ is unbiased while $\breve{\lambda}_j^{\hat{\kappa}}$ in general is not. Since the values of $\breve{\lambda}_j^{\hat{\kappa}}$ are consistently higher than the ones of $\breve{\lambda}_j$ in our numerical computations, this could indicate that $\breve{\lambda}_j^{\hat{\kappa}}$ over-estimates λ_j. We shall therefore in Section 12.3 use the $\breve{\lambda}_j$'s as our estimators of the λ_j's.

There are only small changes in $\hat{\boldsymbol{\beta}}_j$ from Table 12.1 to Table 12.3. Even if there are less variation in the estimates of the ϱ_j's now compared to in Table 12.2, the difference between year 1 and 2 is still considerable. In particular, $\hat{\varrho}_1$ seems rather high. This high

year	$\check{\varphi}_j$ (11.2)	$\check{\lambda}_j$ (10.47)	$\check{\lambda}_j^{\xi}$ (10.31)	$\dot{\boldsymbol{\beta}}_j$ (10.12)	$\hat{\boldsymbol{\beta}}_j$ (10.11)
1	36 864.319	0.0330767	0.0549988	0.4287238 0.0020669	0.4222943 0.0022794
2	70 424.611	0.0253181	0.044204	0.4423509 0.0018885	0.4441854 0.0019285
3	57 011.783	0.0380075	0.0533177	0.4325939 0.0026476	0.4501533 0.0026574

Table 12.5: The original estimators of φ_j, λ_j and $\boldsymbol{\beta}_j$ (liability).

year	1	2
${}_1\hat{\varrho}_j$	0.2815284	0.7929003

${}_1\hat{\pi}_{21}$	${}_1\hat{\pi}_{32}$
0.32	0.65

Table 12.6: Estimators of ϱ_j, π_{21} and π_{32} (liability).

value gives an estimate of π_{21} of 1! This can be caused by the short time-period between the observations of the prices of the car models in year 1 and 2. The same situation occured in Table 12.2. We shall remove the variation in the estimates of the ϱ_j's by assuming that $\varrho_j \equiv \varrho$. We are willing to do this since the observation period of three years is short. Using (10.74) the value of this estimator is ${}_1\hat{\varrho} = 0.88044787$.

Liability insurance

In this subsection we are going to estimate the parameters for (third party) liability insurance.

The estimates analogous to the ones in Table 12.1 are given in Table 12.5. Here the regressors of the estimators of $\boldsymbol{\beta}_j$ are from the top and down: a constant and engine power. The estimates of the $\boldsymbol{\beta}_j$'s are quite stable except for the regressor engine power which increases by 40% from year 2 to year 3. The estimates of the φ_j's vary greatly over the three year period with almost a doubling from year 1 to year 2. This can be explained by the fact that from 1988 on, extra reserves were put on each claim in cases of bodily injuries, to cover future claims due to delayed injuries (e.g. whiplash). Thus, for liability insurance it is not reasonable to assume that the φ_j's are influenced by inflation only and the original estimators will therefore be kept. In Table 12.6 estimates of ϱ_j, π_{21} and π_{32} are given. The estimate of ϱ_1 is very low. This will give little weight to the observation and the prior estimator from year 1 in the empirical recursive credibility estimators, which is unfortunate. The situation improves for the estimate of ϱ_2. The estimate of the correlation π_{21} is also very low.

12.2.2 Estimator of φ_j based on all common car models

To make a comparison with the estimates given in subsection 12.2.1 we will in this subsection give estimates where the ones of φ_j are based on all common car models in the portfolio

year	$\check{\varphi}_j$ (11.2)	$\check{\lambda}_j$ (10.47)	$\check{\lambda}_j^{\hat{\kappa}}$ (10.31)	$\check{\boldsymbol{\beta}}_j$ (10.12)	$\hat{\boldsymbol{\beta}}_j$ (10.11)
1	65 890.249	0.6943819	2.2245343	-0.403245 0.0142767 0.004672	-0.649292 0.0176829 0.0049949
2	69 444.327	0.6926827	2.2184176	-0.594316 0.0123948 0.0063159	-0.773039 0.0153916 0.0062919
3	52 601.357	0.7736816	3.8801143	-0.387578 0.0139309 0.0021017	-0.923543 0.0231861 0.0009204

Table 12.7: The estimators of φ_j, λ_j and $\boldsymbol{\beta}_j$, where the estimators of φ_j are based on common car models from the whole portfolio (vehicle damage).

and with weights $u_{kj}^{\varphi} = (I_{kj} - 1)/(\sum_{r \in \mathcal{A}_j} I_{rj} - K_j)$. By common car models we here mean those car models for which we have a reasonable number of policies in the insurance portfolio, and so on. Whether a car model is common or not is decided by a subjective judgment. In year 1 there were 1310 distinct car models, 1391 in year 2, and 1230 in year 3. Hence we base the estimators on more information than in subsection 12.2.1. The estimators of φ_j based on these weights turned out to be more stable than the ones based on the weights $u_{kj}^{\varphi} = 1/K_j$ (this was also the case for the estimates of φ_j based on observations from a sample of car models, which we have not presented here). The reason for this is that car models with large values of $\check{\varphi}_j$ had a relatively small number of policies associated with them. Car models with a large number of policies associated with them had small values of $\check{\varphi}_j$. By reasons mentioned in the beginning of subsection 12.2.1, it is difficult in practice to base the estimators of the other parameters on all common car models in the portfolio, hence the estimation of $\boldsymbol{\beta}_j$, λ_j and ϱ_j will still be based on a sample of car models from the portfolio.

As in subsection 12.2.1 we exclude car models with zero exposure volume and with one or no policies associated with them.

Vehicle damage insurance

The estimates are given in Table 12.7 and 12.8. We see from Table 12.7 that the estimates do not vary that much from year to year compared to the ones in Table 12.1 (except for $\check{\lambda}_j^{\hat{\kappa}}$, $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\beta}}_3$). In this subsection we will not assume that the estimators of the φ_j's are influenced by inflation only. Hence the estimates given in Table 12.7 will not be changed in the way we did in subsection 12.2.1. Compared to Table 12.3 the estimators of the φ_j's in Table 12.7 are much lower. So in this case the pure random fluctuation of the Y_{kji}'s is estimated lower. In addition, the estimates of the λ_j's are considerably lower in Table 12.3 than in Table 12.7. From (10.47) we see that for $\check{\lambda}_j$ this is a natural consequence of the estimator of φ_j being lower in Table 12.7 than in Table 12.3. The situation is not that simple for $\check{\lambda}_j^{\hat{\kappa}}$ since from (10.31) we have that this estimator depends on both $\check{\varphi}_j$ and $\check{\lambda}_j$ in a rather complicated way, but from the discussion of Table 12.3 we remember that it is not surprising that the difference between $\check{\lambda}_j$ and $\check{\lambda}_j^{\hat{\kappa}}$ is larger in Table 12.7 than in Table 12.3

year	1	2
${}_1\hat{\varrho}_j$	0.5508715	0.3534225

${}_1\hat{\pi}_{21}$	${}_1\hat{\pi}_{32}$
0.55	0.33

Table 12.8: Estimators of ϱ_j, π_{21} and π_{32} where the estimator of φ_j is based on common car models from the whole portfolio (vehicle damage).

year	$\check{\varphi}_j$ (11.2)	$\check{\lambda}_j$ (10.47)	$\check{\lambda}_j^{\hat{\kappa}}$ (10.31)	$\dot{\beta}_j$ (10.12)	$\hat{\beta}_j$ (10.11)
1	30 774.715	0.051415	0.0993959	0.4287238 0.0020669	0.4174625 0.0023873
2	44 694.201	0.0938426	0.2410204	0.4423509 0.0018885	0.4462326 0.0020389
3	46 514.881	0.0697504	0.1179974	0.4325939 0.0026476	0.4565185 0.0026719

Table 12.9: The estimators of φ_j, λ_j and β_j, where the estimators of φ_j are based on common car models from the whole portfolio (liability).

since $\hat{\kappa}_j = \check{\varphi}_j/\check{\lambda}_j$ is much lower in Table 12.7. In the present situation the estimation error of the prior estimators are estimated higher than in Table 12.3. The estimates of the ϱ_j's and the $\pi_{j+1,j}$'s given by Table 12.8 are much lower than the corresponding values given by Table 12.4. This is because the values of $\check{\lambda}_j$ are higher in the present situation while $\hat{\lambda}_{j+1,j}$ is the same. It is therefore reason to question the goodness of the estimators of ϱ_j and π_{ij}. This large difference in the estimates of the parameters will certainly have an impact on the estimated credibility estimators. In the next section we will compare the estimated credibility estimators based on these two sets of estimates.

Liability insurance

The estimates are given in Table 12.9 and 12.10. From the estimators of the φ_j's in Table 12.9 we see very clearly that year 1 is special compared to the other two years. The explanation for why this could be expected was given in subsection 12.2.1. We see the same tendency as in vehicle damage, namely that the φ_j's are estimated lower and the λ_j's higher than in Table 12.5. The $\hat{\beta}_j$'s have only minor changes from Table 12.5 to Table 12.9. The estimates of the ϱ_j's are small in Table 12.10. Compared to Table 12.6 ${}_1\hat{\varrho}_2$ is much lower. This is because of the same reason as in vehicle damage, namely that $\check{\lambda}_j$ is higher here

year	1	2
${}_1\hat{\varrho}_j$	0.181115	0.213919

${}_1\hat{\pi}_{21}$	${}_1\hat{\pi}_{32}$
0.13	0.25

Table 12.10: Estimators of ϱ_j, π_{21} and π_{32} where the estimator of φ_j is based on common car models from the whole portfolio (liability).

and $\hat{\lambda}_{j+1,j}$ is the same. But the effect is even more extreme here than in vehicle damage. This will give little weight to the information lying in the past for the estimation of the credibility estimator of year 4 and thus relatively large weight to the prior estimator in year 4.

12.3 Computing the empirical credibility estimators

In this section we are going to compute the empirical recursive credibility estimators and their estimated estimation errors for a sample of the car models which were in the portfolio in 1989. We have selected every 15 car model of the 380 car models which were in the portfolio in this particular year. We will begin by presenting the technical specifications which we use in $\mathbf{t}_{kj}$, then give the risk volumes and the observations. We are then going to estimate the recursive credibility estimators and their estimation errors for both sets of estimators of parameters presented in Section 12.2. At the end of this section we will compare our estimated credibility estimators (5.8) with the corresponding estimators derived by using the idea behind SUNDT(1987b) on the credibility estimators from the non-hierarchical model in SUNDT(1987a). In order to reduce the number of tables we will only present the computations for vehicle damage covers and skip liability covers.

The technical specifications needed in our computations are given in Table 12.11 for each car model. To save space we have only included specifications valid in 1989. The model code of each car model is grouped into three categories according to type of car: 000–099 are reserved for vans, 100–699 for passenger cars and 700–999 for estate cars.

The risk volumes and observations for the period 1987–1989 are presented in Table 12.12.

The values of the empirical recursive credibility estimators and their estimation errors are given in Table 12.13. In this table we have used the estimators of the structural parameters presented in Tables 12.3 and 12.4. For year 4 we have used $\varrho_3^* = \hat{\varrho}_2$, $\lambda_4^* = \hat{\lambda}_3$, $\boldsymbol{\beta}_4^* = \hat{\boldsymbol{\beta}}_3$, and $\mathbf{t}_{k4} = \mathbf{t}_{k3}$.

Even though Isuzu Gemini has an extreme value of $Y_{k3\cdot}$ the value of $\tilde{m}^*_{k,4|3}$ is close to $\mu^*_{k4}(=0.878)$ since the volume is low. For Toyota Corolla $\hat{\tilde{m}}_{k,2|1}$ and $\hat{\tilde{m}}_{k,3|2}$ are quite close to $Y_{k1\cdot}$ and $Y_{k2\cdot}$, respectively. This is because the volume is large so the weight given to the observation from the last year is relatively large. In year 3 the volume is even larger, but nevertheless $\tilde{m}^*_{k,4|3}$ is not so close to $Y_{k3\cdot}$. The reason for this is that $\hat{\tilde{m}}_{k,3|2}$ contains a lot of information so less weight is given to $Y_{k3\cdot}$. For many of the car models $\tilde{m}^*_{k,4|3}$ is close to $\hat{\tilde{m}}_{k,3|2}$. This is not surprising since we have used the same estimates in year 4 as in year 3. In addition, the weight given to the observation from the last year for all the car models is relatively small because $\hat{\kappa}_3 = \hat{\varphi}_3/\hat{\lambda}_3$ is relatively large.

All the car models except for Suzuki Swift and Mazda 929 have $\hat{\psi}_{k,2|1} < \hat{\psi}_{k,1|0}$. This is so because $\hat{\tilde{m}}_{k,2|1}$ does not base solely on prior information like $\hat{\tilde{m}}_{k,1|0}$ does. The two car models this is not valid for are the ones having the lowest volume in year 1; that is, the information lying in the observations from year 1 is too trifling to be able to compensate for the higher estimate of λ_j in year 2 than in year 1 (a higher value of λ_j means that the prior estimator is less reliable in that year). $\hat{\psi}_{k,3|2}$ is lower than $\hat{\psi}_{k,2|1}$ and for many car models significantly lower. The reason for this is that $\hat{\varrho}_2^2\hat{\lambda}_2 > \hat{\lambda}_3$ and hence $\hat{\psi}_{k,3|2} < \hat{\varrho}_2^2(1-\hat{\zeta}_2)\hat{\psi}_{k,2|1}$. This difference is increasing in the risk volume. Most of the car models has $\psi^*_{k,4|3} < \hat{\psi}_{k,3|2}$, but the difference is not large. For our sample we have $\psi^*_{k,4|3} \leq \hat{\psi}_{k,b_k|b_k-1}$ with equality

Make	Model	Name	Year 3 H.P.	Weight	Price	Price per kilo
14	432	BMW 318 I	113	1 100	246 000	223.636
14	801	BMW 320 I Touring	129	1 250	313 000	250.400
15	557	Citroen CX 2500 GTI	123	1 365	266 000	194.872
16	536	Fiat Croma Turbo	150	1 265	291 000	230.040
18	456	Ford Sierra 2.0I Ghia	120	1 075	269 000	250.233
19	426	Suzuki Swift GTI, GXI	101	805	150 000	186.335
25	505	Mercedes Benz 190 E 2.3-16	195	1 270	619 000	487.402
25	602	Mercedes Benz 300 SE	180	1 525	640 000	419.672
25	905	Mercedes Benz 230 TE	132	1 415	480 000	339.223
31	377	Opel Kadett 1.6 I	75	885	138 000	155.932
33	414	Peugeot 205 CT, CJ	80	880	180 000	204.545
33	855	Peugeot 405 SRI Estate Car	122	1 080	238 000	220.370
37	567	Saab 9000 T16	165	1 335	458 000	343.071
45	413	VW Golf 1.8 CL	90	945	150 000	158.730
46	341	Volvo 460 GL, GLE	102	1 030	206 000	200.000
46	915	Volvo 740 GLT Estate Car	155	1 400	432 000	308.571
51	509	Honda Prelude 2.0I-16 4WS	138	1 045	313 000	299.522
53	815	Subaru 1.8 DL 4WD Estate Car	98	1 120	179 000	159.821
56	302	Isuzu Gemini 1.5 LT, LD	71	890	115 000	129.213
76	403	Plymouth/Chrysler	142	1 590	322 000	202.516
92	552	Alfa Romeo GTV 2.0	130	1 155	208 000	180.087
96	315	Toyota Corolla 1.6 DX, XL	105	985	150 000	152.284
97	345	Nissan Sunny 1.6 SLX	90	1 015	149 000	146.798
98	212	Mazda 121 1.3 LX	66	800	120 000	150.000
98	575	Mazda 929 3.0I GLX V6	170	1 560	350 000	224.359

Table 12.11: The technical specifications in year 3 of each car model included in our study. H.P.=Horse Power (engine power). Weights are given in kilograms and prices in Norwegian kroner.

Make	Model	$w_{k1\cdot}$	$w_{k2\cdot}$	$w_{k3\cdot}$	$Y_{k1\cdot}$	$Y_{k2\cdot}$	$Y_{k3\cdot}$
14	432	358 918	434 936	426 655	2.620	2.975	1.660
14	801	0	819	2 743	·	1.408	0.186
15	557	117 125	108 802	89 375	1.171	1.218	0.898
16	536	53 537	101 076	84 621	5.867	5.509	3.087
18	456	137 730	168 135	179 479	3.228	5.911	2.753
19	426	24 210	33 783	37 254	0	3.580	7.608
25	505	209 930	158 013	124 423	4.282	2.507	4.572
25	602	53 072	81 077	87 848	0.454	3.622	5.664
25	905	66 715	88 993	150 335	3.398	5.646	2.542
31	377	0	0	45 731	·	·	0.996
33	414	52 655	80 723	126 796	0.875	2.086	1.564
33	855	0	0	13 420	·	·	1.530
37	567	130 242	189 515	229 537	1.729	4.841	3.087
45	413	784 315	993 737	1 225 766	1.306	1.527	1.195
46	341	0	0	0	·	·	·
46	915	0	0	0	·	·	·
51	509	0	17 821	27 929	·	13.863	5.405
53	815	72 041	97 534	121 372	0.722	0.966	1.044
56	302	0	0	1 154	·	·	13.867
76	403	0	0	788	·	·	0
92	552	42 933	40 553	40 248	4.718	6.583	4.116
96	315	658 599	1 520 804	1 945 245	1.035	1.259	0.783
97	345	67 106	138 770	178 703	0.958	1.879	1.069
98	212	0	4 619	25 306	·	6.837	1.835
98	575	3 238	27 932	30 740	0	1.456	9.439

Table 12.12: The risk volumes and observations for each car model in the three year period. Missing values are denoted by a dot $(\cdot)$.

Make	Model	$\hat{\tilde{m}}_{k,1\|0}$	$\hat{\tilde{m}}_{k,2\|1}$	$\hat{\tilde{m}}_{k,3\|2}$	$\tilde{m}^*_{k,4\|3}$	$\hat{\psi}_{k,1\|0}$	$\hat{\psi}_{k,2\|1}$	$\hat{\psi}_{k,3\|2}$	$\psi^*_{k,4\|3}$
14	432	2.135	2.363	2.068	1.959	0.313	0.233	0.110	0.125
14	801	·	2.562	2.032	2.026	·	0.330	0.249	0.249
15	557	2.544	2.193	1.524	1.520	0.313	0.286	0.184	0.188
16	536	2.883	2.780	2.902	2.847	0.313	0.308	0.198	0.198
18	456	2.293	2.415	2.664	2.581	0.313	0.280	0.167	0.169
19	426	1.911	1.759	1.491	1.707	0.313	0.319	0.228	0.226
25	505	4.679	4.776	3.062	3.236	0.313	0.262	0.159	0.168
25	602	4.346	4.108	2.832	3.076	0.313	0.308	0.204	0.201
25	905	3.107	3.191	2.536	2.500	0.313	0.303	0.199	0.190
31	377	·	·	0.989	0.989	·	·	0.250	0.239
33	414	1.687	1.602	1.148	1.190	0.313	0.308	0.204	0.196
33	855	·	·	1.868	1.863	·	·	0.250	0.247
37	567	3.781	3.385	2.759	2.807	0.313	0.282	0.163	0.163
45	413	1.630	1.357	1.165	1.181	0.313	0.186	0.066	0.092
46	341	·	·	1.506	1.506	·	·	0.250	0.250
46	915	·	·	2.558	2.558	·	·	0.250	0.250
51	509	·	3.082	2.559	2.606	·	0.330	0.242	0.237
53	815	1.588	1.427	1.236	1.234	0.313	0.301	0.195	0.191
56	302	·	·	0.878	0.894	·	·	0.250	0.249
76	403	·	·	2.165	2.163	·	·	0.250	0.250
92	552	2.403	2.374	2.303	2.326	0.313	0.312	0.220	0.220
96	315	1.489	1.187	1.318	1.179	0.313	0.196	0.052	0.083
97	345	1.440	1.285	1.275	1.242	0.313	0.303	0.185	0.179
98	212	·	·	0.831	0.858	·	·	0.250	0.244
98	575	3.397	3.075	2.576	2.798	0.313	0.329	0.236	0.233

Table 12.13: The empirical recursive credibility estimators and their estimation errors. The estimators of the structural parameters are based on a sample of the car models. (Missing values=·).

Make	Model	$\hat{\mu}_{k1}$	$\hat{\mu}_{k2}$	$\hat{\mu}_{k3}$
14	432	2.135	2.192	1.726
25	505	4.679	4.874	3.516
33	414	1.687	1.666	1.153
45	413	1.630	1.526	1.239
96	315	1.489	1.408	1.474

Table 12.14: The prior estimates in each year.

if and only if $w_{k3\cdot} = 0$. So our credibility estimators have been improved by including observations.

To see what impact the varying values of the $\hat{\beta}_j$'s from Table 12.3 have on the estimates of the μ_{kj}'s, we have computed the values of $\hat{\mu}_{kj}$ $(j = 1,2,3)$ for 5 car models. These values are presented in Table 12.14. For all the five car models the changes from $\hat{\mu}_{k1}$ to $\hat{\mu}_{k2}$ are relatively small. Toyota Corolla is the only car with $\hat{\mu}_{k3} > \hat{\mu}_{k2}$. This is due to a combination of a low price per kilo and a relatively large engine. This car is also the one having the smallest difference (in percentages) between $\hat{\mu}_{k2}$ and $\hat{\mu}_{k3}$ caused by the same reason. Peugeot 205 has the largest difference between $\hat{\mu}_{k2}$ and $\hat{\mu}_{k3}$ since it has a high price per kilo and a relatively small engine. For Mercedes Benz 190 this difference is also large since it has a high price per kilo. But also more "modest" cars like BMW 318 and VW Golf 1.8 have a considerable difference between $\hat{\mu}_{k2}$ and $\hat{\mu}_{k3}$.

In Table 12.15 we have compared the credibility estimators $\tilde{m}_{s,t|t}$ $(t = 1,2,3)$ defined by (5.8) with the estimators resulting from a combination of the non-hierarchical credibility estimators of SUNDT(1987a) and the idea of SUNDT(1987b). This idea is described in Section 7.1. In his numerical example SUNDT(1987a) uses the credibility estimator of $m_{st}(\Theta_{st})$ based on observations from year t only, that is,

$$\tilde{m}_{st} = \frac{v_{st\cdot}\lambda_t}{v_{st\cdot}\lambda_t + \varphi_t} Y_{st\cdot} + \frac{\varphi_t}{v_{st\cdot}\lambda_t + \varphi_t} \mu_{st}.$$

If we use the idea of SUNDT(1987b) on this estimator we shall replace μ_{st} by an old estimator. This leads to the estimators (which are no longer *credibility* estimators):

$$\bar{m}_{s,t|t} = \frac{v_{st\cdot}\lambda_t}{v_{st\cdot}\lambda_t + \varphi_t} Y_{st\cdot} + \frac{\varphi_t}{v_{st\cdot}\lambda_t + \varphi_t} \bar{m}_{s,t-1|t-1}.$$

From (5.8) and (5.11) we have that

$$\tilde{m}_{s,t|t} = \zeta_{st} Y_{st\cdot} + (1 - \zeta_{st}) \left[\varrho_{t-1} \left(\tilde{m}_{s,t-1|t-1} - \mu_{s,t-1}\right) + \mu_{st}\right].$$

Replacing the structural parameters by their estimators we get the empirical estimators which we have compared in Table 12.15. There is clearly larger variation in $\hat{\bar{m}}_{k,t|t}$ compared to $\hat{\tilde{m}}_{k,t|t}$. This is seen from $\sum_k \sum_{t=2}^{3} (\hat{\bar{m}}_{k,t|t} - \hat{\bar{m}}_{k,t-1|t-1})^2 = 6.49$ while $\sum_k \sum_{t=2}^{3} (\hat{\tilde{m}}_{k,t|t} - \hat{\tilde{m}}_{k,t-1|t-1})^2 = 1.95$, where k is summed over the 16 car models which have risk volume greater than zero in all three years. The estimators of Sundt have higher credibility since $\hat{\psi}_{k,t|t-1} \leq \hat{\lambda}_t$ $(t = 1,2,3)$, but this does not lead to a larger fluctuation of $\hat{\tilde{m}}_{k,t|t}$. Our example probably favours the estimators of Sundt since $\hat{\kappa}_3$ is considerably larger than both $\hat{\kappa}_1$ and $\hat{\kappa}_2$. This gives relatively small weight to $Y_{k3\cdot}$ in $\hat{\tilde{m}}_{k,3|3}$ which moderates the possible

Make	Model	$\hat{\tilde{m}}_{k,1\|1}$ $\hat{\tilde{m}}_{k,1\|1}$	$\hat{\tilde{m}}_{k,2\|2}$	$\hat{\tilde{m}}_{k,3\|3}$	$\hat{\tilde{m}}_{k,2\|2}$	$\hat{\tilde{m}}_{k,3\|3}$
14	432	2.330	2.613	2.282	2.581	1.991
14	801	·	·	2.554	2.561	2.025
15	557	2.298	2.121	1.998	2.052	1.477
16	536	3.155	3.517	3.476	3.177	2.916
18	456	2.484	3.281	3.184	3.131	2.675
19	426	1.829	1.929	2.182	1.861	1.740
25	505	4.567	4.111	4.173	4.358	3.198
25	602	3.995	3.947	4.117	4.050	3.065
25	905	3.139	3.486	3.336	3.506	2.537
31	377	·	·	·	·	0.989
33	414	1.615	1.674	1.659	1.660	1.195
33	855	·	·	·	·	1.862
37	567	3.379	3.751	3.603	3.714	2.811
45	413	1.438	1.495	1.313	1.442	1.174
46	341	·	·	·	·	1.506
46	915	·	·	·	·	2.558
51	509	·	·	3.160	3.418	2.652
53	815	1.485	1.407	1.360	1.363	1.216
56	302	·	·	·	·	0.897
76	403	·	·	·	·	2.162
92	552	2.575	2.848	2.908	2.646	2.380
96	315	1.239	1.253	0.920	1.232	1.139
97	345	1.386	1.485	1.409	1.396	1.245
98	212	·	·	·	·	0.862
98	575	3.377	3.285	3.512	2.998	2.817

Table 12.15: Comparison between the estimators of Sundt and our credibility estimators. (Missing values=·).

Make	Model	$\hat{\hat{m}}_{k,1\|0}$	$\hat{\hat{m}}_{k,2\|1}$	$\hat{\hat{m}}_{k,3\|2}$	$\hat{\hat{m}}^*_{k,4\|3}$	$\hat{\psi}_{k,1\|0}$	$\hat{\psi}_{k,2\|1}$	$\hat{\psi}_{k,3\|2}$	$\hat{\psi}^*_{k,4\|3}$
14	432	2.241	2.434	2.107	1.840	0.694	0.526	0.702	0.690
14	801	·	2.689	2.294	2.268	·	0.693	0.773	0.770
15	557	2.707	2.096	1.795	1.822	0.694	0.576	0.725	0.718
16	536	3.083	3.286	3.333	2.920	0.694	0.617	0.728	0.719
18	456	2.418	2.611	2.858	2.334	0.694	0.568	0.717	0.703
19	426	1.999	1.685	1.672	2.346	0.694	0.650	0.749	0.738
25	505	5.017	4.798	3.499	4.091	0.694	0.548	0.718	0.710
25	602	4.667	3.807	3.315	3.979	0.694	0.617	0.732	0.718
25	905	3.302	3.312	2.818	2.514	0.694	0.606	0.730	0.707
31	377	·	·	0.959	0.964	·	·	0.774	0.735
33	414	1.742	1.525	1.142	1.223	0.694	0.617	0.732	0.710
33	855	·	·	2.108	2.074	·	·	0.774	0.758
37	567	4.054	3.206	3.308	3.191	0.694	0.571	0.715	0.699
45	413	1.690	1.396	1.283	1.271	0.694	0.505	0.695	0.682
46	341	·	·	1.626	1.626	·	·	0.774	0.774
46	915	·	·	2.954	2.954	·	·	0.774	0.774
51	509	·	3.214	3.120	2.985	·	0.693	0.761	0.745
53	815	1.645	1.356	1.355	1.377	0.694	0.602	0.728	0.711
56	302	·	·	0.842	0.918	·	·	0.774	0.772
76	403	·	·	2.555	2.545	·	·	0.774	0.773
92	552	2.549	2.733	2.753	2.607	0.694	0.627	0.744	0.736
96	315	1.537	1.217	1.579	1.355	0.694	0.509	0.692	0.680
97	345	1.483	1.261	1.375	1.249	0.694	0.605	0.721	0.703
98	212	·	·	0.745	0.849	·	·	0.774	0.747
98	575	3.661	3.291	3.061	3.856	0.694	0.686	0.754	0.742

Table 12.16: The values of the empirical recursive credibility estimators and their estimation errors. The estimators of the φ_j's are based on all common car models from the whole of the portfolio. (Missing values=·).

large values of $Y_{k3\cdot}$. We must also remember that in SUNDT(1987a) the risk level of each car model is assumed to be independent of time. This is in contrast to our models which allow the risk levels to vary from year to year. Then it is perhaps not unreasonable that $\hat{\hat{m}}_{k,t|t}$ fluctuates less than $\hat{m}_{k,t|t}$.

In Table 12.16 the same quantities are presented as in Table 12.13, but the estimates of the φ_j's are based on all common car models from the whole portfolio. We will compare the values of this table with the values of Table 12.13. Since the values of $\hat{\kappa}_t$ $(t = 1, 2, 3)$ are much lower here than the corresponding estimates in Table 12.13, $\hat{\zeta}_{kt}$ will become much larger here. Even if the estimates of ϱ_t is lower here than in Table 12.13 the weight given to the observation from the last year when computing $\hat{\hat{m}}_{k,t|t-1}$ $(t = 1, 2, 3)$ and $\hat{\hat{m}}^*_{k,4|3}$ is much larger here. The difference between the two tables is smallest for car models with the largest volume since $\hat{\kappa}_t$ has less influence on $\hat{\zeta}_{kt}$, and largest for the ones with little volume. The weight given to $\hat{\hat{m}}_{k,t|t-1}$ when computing $\hat{\hat{m}}_{k,t+1|t}$ therefore becomes considerably lower here than in Table 12.13. The estimators of the structural parameters used in Table 12.16 yield

		$\hat{\tilde{m}}_{k,1\|1}$				
Make	Model	$\hat{\tilde{m}}_{k,1\|1}$	$\hat{\tilde{m}}_{k,2\|2}$	$\hat{\tilde{m}}_{k,3\|3}$	$\hat{\tilde{m}}_{k,2\|2}$	$\hat{\tilde{m}}_{k,3\|3}$
14	432	2.541	2.893	1.829	2.849	1.726
14	801	·	·	2.592	2.679	2.213
15	557	1.858	1.525	1.169	1.679	1.300
16	536	4.087	4.801	3.850	4.338	3.200
18	456	2.897	4.785	3.311	4.522	2.783
19	426	1.593	2.094	4.046	2.140	3.729
25	505	4.511	3.285	4.117	3.527	4.174
25	602	3.156	3.364	4.661	3.730	4.607
25	905	3.341	4.425	3.128	4.332	2.631
31	377	·	·	·	·	0.974
33	414	1.432	1.724	1.620	1.759	1.412
33	855	·	·	·	·	2.013
37	567	2.709	4.104	3.319	4.202	3.141
45	413	1.348	1.510	1.212	1.511	1.200
46	341	·	·	·	·	1.626
46	915	·	·	·	·	2.954
51	509	·	·	3.852	4.821	3.777
53	815	1.247	1.109	1.067	1.178	1.160
56	302	·	·	·	·	1.059
76	403	·	·	·	·	2.526
92	552	3.224	4.192	4.164	3.765	3.248
96	315	1.098	1.249	0.799	1.255	0.813
97	345	1.265	1.621	1.221	1.599	1.157
98	212	·	·	·	·	1.041
98	575	3.541	3.086	5.064	2.894	5.012

Table 12.17: Comparison between the estimators of Sundt and our credibility estimators. (Missing values=·).

higher prior estimators in all years for nearly all car models. Much of the same analysis as in Table 12.13 applies to Table 12.16 as well. But the difference between $\hat{\tilde{m}}^*_{k,4|3}$ and $\hat{\tilde{m}}_{k,3|2}$ is larger in Table 12.16. This is because the weight given to $Y_{k3\cdot}$ when computing $\hat{\tilde{m}}^*_{k,4|3}$ is larger in Table 12.16 than in Table 12.13. $\hat{\psi}_{k,3|2}$ is larger than both $\hat{\psi}_{k,2|1}$ and $\hat{\psi}^*_{k,4|3}$ for all car models. $\hat{\psi}_{k,2|1}$ is smaller than $\hat{\psi}_{k,1|0}$ because $\hat{\lambda}_2 < \hat{\lambda}_1$. The difference is largest for those car models with the largest volume in year 1. $\hat{\psi}^*_{k,4|3}$ is larger than $\hat{\psi}_{k,1|0}$ except for BMW 318, VW Golf and Toyota Corolla. These three car models have all a large risk volume throughout the observation period and two of them have an increasing risk volume. The other car models which are best off (that is, the ones with the smallest increase in estimation error from year 1 to year 4) are the ones with an increasing risk volume. We also notice that $\hat{\psi}_{k,3|2}$ is close to $\check{\lambda}_3$ for all the car models.

Table 12.17 is the same as Table 12.15 but the estimates of the φ_t's are based on all common car model from the whole portfolio. In this case we have $\sum_k \sum_{t=2}^{3}(\hat{\tilde{m}}_{k,t|t} - \hat{\tilde{m}}_{k,t-1|t-1})^2 = 27.43$ and $\sum_k \sum_{t=2}^{3}(\hat{\tilde{m}}_{k,t|t} - \hat{\tilde{m}}_{k,t-1|t-1})^2 = 27.84$ where the sum still is taken over the 16 car models which have $w_{kt.} > 0$ $(t = 1, 2, 3)$. So there is no significant difference between the two estimators with respect to fluctuation. This similarity can be explained by the fact that since $\hat{\psi}_{k,3|2}$ is close to $\check{\lambda}_3$ for all car models, the weight given to $Y_{k3.}$ in $\hat{\bar{m}}_{k,3|3}$ and $\hat{\bar{m}}_{k,3|3}$ is close to each other. The weight given to $Y_{k2.}$ in $\hat{\bar{m}}_{k,2|2}$ and $\hat{\bar{m}}_{k,2|2}$ is also quite similar.

When comparing the estimator of Sundt with our credibility estimator we have to bear in mind that our estimator is the best among the linear estimators under our model assumptions (when using expected quadratic loss as optimality criterion). So if our model assumptions are completely satisfied in practice the estimator of Sundt can not be better than ours. However, if one feels that our model assumptions are not completely realistic the estimator of Sundt could be a better estimator than ours. This is perhaps in particular the case if one is apt to use another optimality criterion. The accuracy of the two estimators is difficult to compare since the estimation errors of $\bar{m}_{k,t|t}$ $(t = 1, 2, 3)$ are not easy to calculate.

Chapter 13

Discussion

When constructing a model we have to take two things into consideration: mathematical simplicity and realism. Unfortunately it seems that we are not able to fulfil both the requirements to their utmost extent. We have to compromise. The model should not make the computations too complicated and it should describe the main features of the process under consideration. This must be our starting point when criticizing a model.

Assumptions 4.1–4.4, 5.2, 5.5, and the assumptions that the structural parameters are independent of the car model are mainly to ensure mathematical simplicity. In particular, Assumption 5.6 makes it easy to compute the credibility estimator.

The object of Assumptions 5.1, 5.3 and 5.7 is mainly to make the models realistic.

But assumptions chosen to make the computations simple were also chosen because they were not unrealistic, and vice versa.

The realism of the assumption of independence between car models depends on the definition of a car model. For instance, it should be obvious that we can learn something about BMW 320 by observing BMW 325. Whereas observing a Lada will probably give us very limited information about a Mercedes Benz. We should therefore be careful when making a division of the portfolio into car models. In Chapter 12 we have defined each variant of a car model as a separate car model. This clearly violates the independence assumption. A natural way to compensate for this is to consider different variants of a car model as one car model. This is what we will do in Section 14.6.

Since traffic intensity, road conditions, the properties of a car model etc. vary from year to year, it is desirable that the structural parameters are allowed to vary as well. In particular the correlation between risk levels (and claim data) should decrease when the distance between the years increases. That is, the importance of the claim data from a specific year should decrease the older it gets. But this is exactly what (5.4) expresses.

In Section 6.3 we saw that the recursive credibility estimators derived under the time-homogeneous model have some attractive properties. These properties may also be valid for the recursive credibility estimators in the models of Chapters 5 and 7, but this depends on the value of the parameters in each particular case.

In Chapter 7 we have proposed a model which takes car models with unknown prices into consideration. It should be interesting to check how this model behaves in practice. This would have been a more elaborate project and was therefore not included in the present study.

We showed in Chapter 8 that the credibility estimators in our models are nothing else than the Kalman filter. But we have done more than just using the Kalman filter. We have stated models with desirable properties and under these models we could reduce the estimation problem considerably by a simple application of linear sufficiency. This made the calculation of the recursive credibility estimators much easier which is very important in practice; instead of having to store the claim data from all of the individual policies in the whole portfolio from each year, we only have to take a weighted average over the policies associated with the car model in question and store that quantity for each year.

In our application of the model in Chapter 5 we used sports cars, cars with four-wheel drive, diesel cars and so on (we only excluded the most expensive cars in addition to cars with zero risk volume and the ones with one or no policies associated with them) for the estimation of $\boldsymbol{\beta}_j$, λ_j and ϱ_j. This is dangerous to do because these rather extreme cars may influence too much on the estimates. This could also be a reason for the instability in our estimates. We included these cars to avoid having too few car models to base the estimation on. This was particularly a problem for 1989. In this year we based the estimation on only 298 car models since many of the car models in this year had volume equal to zero. If we had excluded the above-mentioned cars as well, we would have ended up with approximately 220 car models. This is a substantial reduction from the initial 380 distinct car models.

As already mentioned the estimates presented in Chapter 12 were quite unstable. This instability could be originated from causes other than extreme car models. In fact there has been a considerable change in the composition of the risks in the portfolio during the period from 1987 to 1989. In 1987 the economy in Norway was tighter. This has affected peoples behaviour in the subsequent years. The number of passenger cars sold has decreased, the number of distinct car models to choose from has been reduced, people are buying smaller and less expensive cars, fewer new policies have been taken out, there is an increased tendency to drop the vehicle damage cover (liability cover is compulsory in Norway), the number of older cars has increased, and the mileage has decreased. This has undoubtedly had an influence on the structural parameters. This could be an argument against the way we tried to reduce the variation in the estimates by assuming constantness of the ϱ_j's and that the φ_j's are influenced by inflation only. This was of course a simplification of reality, and we were only willing to make these rather strict assumptions because the observational period was relatively short. However, the fluctuation in the estimates seemed rather extreme and should be amended for in some way or other. There could of course exist alternatives to the way we tried to solve this problem. But there is an argument which supports the way we tried to eliminate the fluctuation in the ${}_1\hat{\varrho}_j$'s. In Chapter 12 we used the technical variables of the car models in 1987 from a list dated November the same year, in 1988 from February 1988, but in 1989 we used a list dated May 1990. The period between the two first lists is only three months while the period between the second and the third list is 27 months! Then it is perhaps not surprising that the correlation between 1987 and 1988 initially was estimated very high and the correlation between 1988 and 1989 was estimated very low. So, some kind of an average between ${}_1\hat{\varrho}_1$ and ${}_1\hat{\varrho}_2$ as an estimator of the ϱ_j's in that time period is maybe not that unrealistic even if one feels that the ϱ_j's have not been constant during this period. The long time-period could also explain the considerable drop between the estimators of $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$.

Before making a final conclusion about the performance of our model in Chapter 5 we should apply the model on real data where the prices for each car model are observed at equally spaced time intervals.

On how many years of observations should we base the estimation of ϱ_j^p? That is, how large should we choose d to be? Before making a decision about this we have to be aware of the fact that the number of car models on which to base the computation of $\hat{\lambda}^p_{ij}$ will roughly be decreasing in $|i-j|$. Hence, it is not necessarily desirable to choose the number of years d as large as possible. One should in each particular situation investigate whether the amount of information gained when increasing d justifies the more elaborate task of handling observations from an increasing number of years.

The estimation of parameters in our models is clearly a topic of further research. As mentioned before we have not tried to derive optimal estimators, merely reasonable estimators which are not too complicated for practical purposes. One may question whether the estimators of ϱ^p_j proposed in this paper are sufficiently simple. They should be practicable for d small (as in our example where $d=1$) but for a large d the estimation process becomes bothersome. If one feels that one ought to base the estimation on observations from a large number of years then one could think over whether one should seek alternative estimators of the ϱ^p_j's. In Chapter 12 we saw that another drawback of the estimators of ϱ^p_j derived in this paper, is that their values decreased if the estimators of the λ^p_j's increased and one could perhaps argue that they are too sensitive towards this. One could perhaps amend for this by basing the estimator of λ_{ij} $(i>j)$ on $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\beta}}_i$ instead of on $\dot{\boldsymbol{\beta}}_j$ and $\dot{\boldsymbol{\beta}}_i$.

As we remember from Section 11.2 we were not quite satisfied with the estimator (11.2) and we therefore modified it. As pointed out in subsection 12.2.1 it was not reasonable to apply this modification on our data material. We are therefore not quite happy with the Poisson-assumption and we should seek alternative ways to estimate the φ_j's.

We must be aware of the fact that our models have their limitations. They do not describe perfectly the processes under consideration. The estimate that our models produces should therefore be combined with an expert's opinion to reach a final estimate. This expert will use his knowledge of the different car models to correct the proposed estimates of the risk level given by the models. He may also take a glimpse at the rating of the car models in competing companies. In this way the estimate produced by our models will serve as a guideline for the expert to make a final classification. For a more thorough discussion on this subject we refer to SUNDT(1987a, subsection 3.7) and SUNDT(1991b, Section 8.2).

We have in this paper derived models for classification of cars, and at least one of the models turned out to be applicable in practice (after some modifications, see Chapter 12). However, the present author feels that there is still some work to be done concerning the estimation of parameters before we have a complete rating system at hand. If one wants to implement the system with the estimators derived in this paper, it should be used with care. The implementation and later updating should be controlled by a statistician (actuary) who should detect and try to modify estimators with unreasonable behaviour.

Chapter 14

Areas of Further Research

14.1 Introduction

During the present study many problems arose which demanded solutions. The work connected with finding solutions to these problems was rather time-consuming and the consequence was that topics which we initially wanted to treat had to be omitted. These topics will be discussed and ways to extend our models will be indicated in this chapter. We will only give suggestions and not try to give any final solutions to the problems. Hence this chapter will be of a summary nature.

14.2 Choice of regressors

In this study we have used the technical variables: a constant term, engine power, and price divided by weight in vehicle damage, and a constant term and engine power in liability insurance. Now and again one should examine whether the regressors currently in use still are able to give a sufficiently good prior estimator of the risk level in that particular year. This is particularly important for new car models for which we do not have any claim experience. If there exists another set of technical variables yielding better prior estimators and possessing other desirable properties (for a discussion on how to select the regressors and which properties they should have, we refer to SUNDT(1987a, subsection 3.5); see also Section 14.3) one should use these variables instead of the set currently in use.

In particular, if we are going to make use of the time-heterogeneous model for two portfolios we have to find a new set of regressors in vehicle damage for the car models in sub-portfolio B.

14.3 Sports cars and other extreme cars

It has been advocated by practitioners in UNI Storebrand that the technical variables we have used do not give an adequate prior estimator for the very expensive cars, sports cars, cars with four-wheel drive, and diesel cars. As proposed by SUNDT(1987a, subsection 3.5) for the two latter, one could include (0,1)-variables for these characteristics. But one has to be careful not to include too many regressors. The regressors should secure monotonocity in the prior estimator, that is, a sports car should be rated higher than an ordinary car, a car with a larger engine power should be rated higher than a car with a smaller engine and so on.

Another way to solve this problem is by dividing the portfolio into sub-portfolios. One sub-portfolio consists of very expensive cars, another of sports cars and so on. The whole analysis described in this paper should then be performed separately on each of the sub-portfolios. A problem with this approach is that, of course, each sub-portfolio will contain a smaller number of car models than the original portfolio. This will make it even more difficult to obtain estimators of the parameters with an acceptable accuracy.

In either case we have the problem on how to distinguish between sports cars and ordinary cars. On which criteria should we base this distinction? What do we mean by a sports car? INGENBLEEK & LEMAIRE(1988) try to shed some light on these problems, but they do not base their analysis on claim data, only on technical characteristics.

14.4 Risk parameter describing the whole portfolio

In some years we may expect more claims and/or larger claims than in other years. This could be a result of varying weather conditions from year to year. If, for instance, the roads are very icy one winter we would expect more claims than is normal from all the car models in the portfolio. We have not taken this into consideration in our models; we have assumed the claim amounts from different car models in the same year to be independent. One way to make up for this deficiency is to modify our models in this paper by including a random risk characteristic for each year describing how this year differs from other years. We may denote these risk characteristics by H_j (capital Greek η) and assume that the H_j's are mutually independent and independent of the Θ_k's. Assumptions 4.1 and 4.2 may now be modified by being valid when the H_j's are given. Assumptions 4.3 and 4.4 are kept unchanged. The assumptions given in Chapters 5–7 should be modified in a similar way.

A more thorough discussion on this subject is given in SUNDT(1979).

14.5 Hierarchical models

In our application of the model in Chapter 5 we considered each variant as a separate car model. In our models we have assumed independence between car models. As pointed out in Chapter 13 this is not realistic for all the car models. We may modify our models to take this into consideration. As proposed by SUNDT(1987a, p. 64) this may be done by introducing hierarchical models on two or three levels. In a model with three levels one could have one level for make of car, one level for model and one level for variant. In a model with two levels we could have one level for make and one for model or one level for model and one for variant.

Another approach to the solution of this problem is given in the next section.

14.6 Joint car model

Since different variants of a car model probably will be dependent of each other we could reduce the dependency between the variants by handling different variants as one car model. Variants which differ considerably from the other variants of the car model should not be pooled together with the other variants. The technical variables for this "car model" could be a weighted average of the technical variables of the variants which make up the car model (the weight being e.g. the risk volume). The technical variables for each of these variants will be incorporated into the (hidden) random risk characteristics of that car model. In this way the total number of car models in the portfolio will be reduced considerably, but there will be more information to base the estimation of the risk level of each car model on.

This approach has also been discussed by SUNDT(1987a, p. 64).

References

BÜHLMANN, H. & STRAUB, E. (1970): Glaubwürdigkeit für Schadensätze, *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker* **70**, 111–133.

DEJONG, P. & ZEHNWIRTH, B. (1983): Credibility theory and the Kalman filter, *Insurance: Mathematics and Economics* **2**, 281–286.

DE VYLDER, F. (1981): Practical credibility theory with emphasis on optimal parameter estimation, *ASTIN Bulletin* **12**, 115–131.

GERBER, H.U. & JONES, D.A. (1975): Credibility formulas of the updating type, *Transactions of the Society of Actuaries* **27**, 31–46.

HESSELAGER, O. (1989): Supplement til Bjørn Sundt: "An introduction to non-life insurance mathematics"; kap 6. In Danish. Unpublished.

INGENBLEEK, J. & LEMAIRE, J. (1988): What is a sports car?, *ASTIN Bulletin* **18**, 175–187.

KARLIN, S. & TAYLOR, H.M. (1975): *A first course in stochastic processes* (2nd ed.), Academic Press Inc., New York.

NEUHAUS, W. (1989): Dynamic linear models and the discrete Kalman filter, *Statistical Memoirs* 1989-1, Institute of Mathematics, University of Oslo.

NORBERG, R. (1980): Emprical Bayes credibility, *Scandinavian Actuarial Journal*, 177–194.

NORBERG, R. (1982): On optimal parameter estimation in credibility, *Insurance: Mathematics and Economics* **1**, 73–89.

OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1987): Bilpriser 1987 – November. Norsk Veiforening, Oslo.

OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1988): Bilpriser 1988 – Februar. Norsk Veiforening, Oslo.

OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1990): Bilpriser 1990 – Mai. Norsk Veiforening, Oslo.

SCHWARTZ, J.T. (1961): *Introduction to matrices and vectors*, McGraw-Hill, New York.

SUNDT, B. (1979): An insurance model with collective seasonal random factors, *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker* **79**, 57–64.

SUNDT, B. (1981): Recursive credibility estimation, *Scandinavian Actuarial Journal*, 3–21.

SUNDT, B. (1983): On time-heterogeneous credibility estimation, *Scandinavian Actuarial Journal*, 183–190.

SUNDT, B. (1987a): Two credibility regression approaches for the classification of passenger cars in a multiplicative tariff, *ASTIN Bulletin* **17**, 41–70.

SUNDT, B. (1987b): Credibility and old estimates, *Insurance: Mathematics and Economics* **6**, 189–194.

SUNDT, B. (1991a): Linearly sufficient fun with credibility, *Insurance: Mathematics and Economics* **10**, 69–74.

SUNDT, B. (1991b): *An introduction to non-life insurance mathematics* (2nd ed.), Verlag Versicherungswirtschaft e.V., Karlsruhe.

WITTING, T. (1986): Über Kredibilitetschätzungen von reinen IBNR-Schäden, *Dissertation* **8042**, Swiss Federal Institute of Technology, Zurich.

WITTING, T. (1987): The linear Markov property in credibility theory, *ASTIN Bulletin* **17**, 71–84.

Address for correspondence:

Arne Eyland
The Wyatt Company A.S
P.O.Box 1508 Vika
N-0117 Oslo
Norway

Appendix A

Useful Identities

For the sake of simplicity we will use the notation from the model of Chapter 5 in this appendix. The identities may be extended in the obvious way to be valid in the model of Chapter 7 as well.

A.1 Covariances

In the present section we will derive some identities concerning unconditional and conditional covariance between observations and risk levels.

In Assumption 5.3 we defined v_{kji} to be identical to w_{kji}. In Section 11.2 we redefined v_{kji} to be $v_{kji} = w_{kji}/e_{kj}$, where e_{kj}=engine power for car model k in year j. This redefinition was performed in order to improve the realism of our models. In either case we can write

$$v_{kji} = h_{kj} w_{kji}, \tag{A.1}$$

with $h_{kj} \equiv 1$ in the former case and $h_{kj} = 1/e_{kj}$ in the latter. Thus we have

$$\operatorname{Var}(Y_{kji}|\Theta_{kj}) = s_j^2(\Theta_{kj})/v_{kji} = s_j^2(\Theta_{kj})/(h_{kj} w_{kji}). \tag{A.2}$$

Summing both sides of (A.1) over i, we get

$$v_{kj\cdot} = h_{kj} w_{kj\cdot}. \tag{A.3}$$

These identities will be utilized in the expressions below.

From Assumptions 4.2, 4.3 and 5.3 we get

$$\operatorname{Cov}(Y_{kji}, Y_{kj'i'}|\Theta_k) = \delta_{j,j'}\delta_{i,i'}\frac{s_j^2(\Theta_{kj})}{v_{kji}}. \tag{A.4}$$

We have

$$\begin{aligned}
\operatorname{Cov}(Y_{kj\cdot}, Y_{kj'i'}|\Theta_k) &\overset{(1)}{=} \frac{1}{w_{kj\cdot}}\sum_{i=1}^{I_{kj}} w_{kji}\operatorname{Cov}(Y_{kji}, Y_{kj'i'}|\Theta_k) \overset{(2)}{=} \frac{1}{w_{kj\cdot}}\sum_{i=1}^{I_{kj}} w_{kji}\delta_{j,j'}\delta_{i,i'}\frac{s_j^2(\Theta_{kj})}{v_{kji}} \\
&\overset{(3)}{=} \frac{1}{w_{kj\cdot}}\sum_{i=1}^{I_{kj}} w_{kji}\delta_{j,j'}\delta_{i,i'}\frac{s_j^2(\Theta_{kj})}{h_{kj} w_{kji}} = \delta_{j,j'}\frac{s_j^2(\Theta_{kj})}{h_{kj} w_{kj\cdot}}.
\end{aligned}$$

where (1) is a consequence of the definition of $Y_{kj\cdot}$, (2) of (A.4), and (3) is a result of (A.1). Using (A.3) we get

$$\mathrm{Cov}(Y_{kj\cdot}, Y_{kj'i'}|\Theta_k) = \delta_{j,j'}\frac{s_j^2(\Theta_{kj})}{v_{kj\cdot}}. \tag{A.5}$$

Further, we have

$$\mathrm{Cov}(Y_{kj\cdot}, Y_{kj'\cdot}|\Theta_k) \overset{(1)}{=} \frac{1}{w_{kj'\cdot}}\sum_{i'=1}^{I_{kj'}} w_{kj'i'}\mathrm{Cov}(Y_{kj\cdot}, Y_{kj'i'}|\Theta_k) \overset{(2)}{=} \frac{1}{w_{kj'\cdot}}\sum_{i'=1}^{I_{kj'}} w_{kj'i'}\delta_{j,j'}\frac{s_j^2(\Theta_{kj})}{v_{kj\cdot}}$$

where (1) follows from the definition of $Y_{kj'\cdot}$ and (2) is a consequence of (A.5). This yields

$$\mathrm{Cov}(Y_{kj\cdot}, Y_{kj'\cdot}|\Theta_k) = \delta_{j,j'}\frac{s_j^2(\Theta_{kj})}{v_{kj\cdot}}. \tag{A.6}$$

Before continuing we remind about the useful identity

$$\mathrm{Cov}(\mathbf{V}, \mathbf{W}') = \mathrm{E}[\mathrm{Cov}(\mathbf{V}, \mathbf{W}'|\mathbf{Z})] + \mathrm{Cov}(\mathrm{E}[\mathbf{V}|\mathbf{Z}], \mathrm{E}[\mathbf{W}'|\mathbf{Z}]) \tag{A.7}$$

valid for arbitrary random vectors $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{Z}$.

We now consider

$$\begin{aligned}
\mathrm{Cov}(Y_{kji}, Y_{k'j'i'}) &\overset{(1)}{=} \delta_{k,k'}\mathrm{Cov}(Y_{kji}, Y_{kj'i'}) \\
&\overset{(2)}{=} \delta_{k,k'}\left\{\mathrm{E}[\mathrm{Cov}(Y_{kji}, Y_{kj'i'}|\Theta_k)] + \mathrm{Cov}(\mathrm{E}[Y_{kji}|\Theta_k], \mathrm{E}[Y_{kj'i'}|\Theta_k])\right\} \\
&\overset{(3)}{=} \delta_{k,k'}\left\{\mathrm{E}[\delta_{j,j'}\delta_{i,i'}\frac{s_j^2(\Theta_{kj})}{v_{kji}}] + \mathrm{Cov}(m_{kj}(\Theta_{kj}), m_{kj'}(\Theta_{kj'}))\right\}
\end{aligned}$$

where (1) follows from Assumption 4.1, (2) from (A.7), and (3) from (A.4), Assumptions 4.3 and 5.1. Using Assumptions 5.4 and 5.5 we have

$$\mathrm{Cov}(Y_{kji}, Y_{k'j'i'}) = \delta_{k,k'}\left\{\delta_{j,j'}\delta_{i,i'}\frac{\varphi_j}{v_{kji}} + \lambda_{j,j'}\right\}.$$

We have from Assumption 4.1

$$\mathrm{Cov}(Y_{kj\cdot}, Y_{k'j'i'}) = \delta_{k,k'}\mathrm{Cov}(Y_{kj\cdot}, Y_{kj'i'}).$$

(A.7) and (A.5) yield

$$\mathrm{Cov}(Y_{kj\cdot}, Y_{k'j'i'}) = \delta_{k,k'}\left[\delta_{j,j'}\frac{\varphi_j}{v_{kj\cdot}} + \lambda_{j,j'}\right]. \tag{A.8}$$

Further from Assumption 4.1 we get

$$\mathrm{Cov}(Y_{kj\cdot}, Y_{k'j'\cdot}) = \delta_{k,k'}\mathrm{Cov}(Y_{kj\cdot}, Y_{kj'\cdot})$$

and using (A.7) and (A.6) we obtain

$$\mathrm{Cov}(Y_{kj\cdot}, Y_{k'j'\cdot}) = \delta_{k,k'}\left[\delta_{j,j'}\frac{\varphi_j}{v_{kj\cdot}} + \lambda_{j,j'}\right]. \tag{A.9}$$

We also have

$$\begin{aligned}\mathrm{Cov}(m_{kj}(\Theta_{kj}), Y_{k'j'i'}) &\overset{(1)}{=} \mathrm{E}[\mathrm{Cov}(m_{kj}(\Theta_{kj}), Y_{k'j'i'}|\Theta_{kj}, \Theta_{k'j'})] \\ &\quad + \mathrm{Cov}(\mathrm{E}[m_{kj}(\Theta_{kj})|\Theta_{kj}, \Theta_{k'j'}], \mathrm{E}[Y_{k'j'i'}|\Theta_{kj}, \Theta_{k'j'}]) \\ &\overset{(2)}{=} \mathrm{Cov}(m_{kj}(\Theta_{kj}), m_{k'j'}(\Theta_{k'j'})) \overset{(3)}{=} \delta_{k,k'}\mathrm{Cov}(m_{kj}(\Theta_{kj}), m_{kj'}(\Theta_{kj'}))\end{aligned}$$

where (1) follows from (A.7), (2) from Assumptions 4.3 and 5.1 and the fact that $m_{kj}(\Theta_{kj})$ is non-random when Θ_{kj} is given, and (3) from Assumption 4.4. Using Assumption 5.5 we get

$$\mathrm{Cov}(m_{kj}(\Theta_{kj}), Y_{k'j'i'}) = \delta_{k,k'}\lambda_{j,j'}. \tag{A.10}$$

Finally,

$$\mathrm{Cov}(m_{kj}(\Theta_{kj}), Y_{k'j'\cdot}) \overset{(1)}{=} \frac{1}{w_{k'j'\cdot}} \sum_{i'=1}^{I_{k'j'}} w_{k'j'i'}\mathrm{Cov}(m_{kj}(\Theta_{kj}), Y_{k'j'i'}) \overset{(2)}{=} \frac{1}{w_{k'j'\cdot}} \sum_{i'=1}^{I_{k'j'}} w_{k'j'i'}\delta_{k,k'}\lambda_{j,j'}$$

where (1) follows from the definition of $Y_{k'j'\cdot}$ and (2) from (A.10). Thus

$$\mathrm{Cov}(m_{kj}(\Theta_{kj}), Y_{k'j'\cdot}) = \delta_{k,k'}\lambda_{j,j'}. \tag{A.11}$$

A.2 Correlations

We assume $(i \geq j)$. The definition of correlation and Assumptions 5.5 and 5.7 yield

$$\pi_{i+1,j} = \frac{\lambda_{i+1,j}}{\sqrt{\lambda_{i+1}\lambda_j}}. \tag{A.12}$$

Inserting Assumption 5.6 into (A.12) and using the definition of correlation and Assumption 5.7 once again gives

$$\pi_{i+1,j} = \frac{\varrho_i\lambda_{ij}}{\sqrt{\lambda_{i+1}\lambda_j}} = \varrho_i\sqrt{\frac{\lambda_i}{\lambda_{i+1}}}\frac{\lambda_{ij}}{\sqrt{\lambda_i\lambda_j}} = \pi_{i+1,i}\pi_{ij}. \tag{A.13}$$

We have from (A.13) that the special case $i = j$ yields (note that $\pi_{ii} = 1\ \forall i$)

$$\pi_{j+1,j} = \varrho_j\sqrt{\frac{\lambda_j}{\lambda_{j+1}}}. \tag{A.14}$$

A.3 Expectations

In the present section we will derive some properties about conditional and unconditional expectations. First we note that

$$\mathrm{E}(V) = \mathrm{E}[\mathrm{E}(V|W)] \tag{A.15}$$

is valid for any pair of random variables V and W.

Consider

$$\mathrm{E}(Y_{kj\cdot}|\Theta_{kj}) \overset{(1)}{=} \frac{1}{w_{kj\cdot}} \sum_{i=1}^{I_{kj}} w_{kji}\mathrm{E}(Y_{kji}|\Theta_{kj}) \overset{(2)}{=} \frac{1}{w_{kj\cdot}} \sum_{i=1}^{I_{kj}} w_{kji} m_{kj}(\Theta_{kj})$$

where (1) follows from the definition of $Y_{kj\cdot}$ and (2) from Assumption 5.1. This yields

$$\mathrm{E}(Y_{kj\cdot}|\Theta_{kj}) = m_{kj}(\Theta_{kj}). \tag{A.16}$$

On the other hand using (A.15) and (A.16) we have

$$\mathrm{E}(Y_{kj\cdot}) = \mathrm{E}[\mathrm{E}(Y_{kj\cdot}|\Theta_{kj})] = \mathrm{E}[m_{kj}(\Theta_{kj})]$$

and by Assumption 5.2 we obtain

$$\mathrm{E}(Y_{kj\cdot}) = \mu_{kj}.$$

Index